

SEQUENCE LISTING

<110> MBARI

DeLong, Edward

Beja, Oded

<120> Light-driven energy generation using proteorhodopsin

<130> MBA-101

<150> 60/201,602

<151> 2000-05-03

<160> 65

<170> PatentIn version 3.0

<210> 1

<211> 105184

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS (complement)

acacccat	ttt	tataaaaaac	ttattgcaga	atcagg	tttt	actccctcat	ggaact	tttaa	480
taaatact	ta	atctcaaaag	agggcaaggt	tgtatccaca	tatggatcaa	aggtaaagcc			540
tgattcaaaa		gagcttatat	cagctataga	aggcttgctg	taaaattatt	acttagaaac			600
taatacag	tt	ttaggcttgt	ttgctgcaaa	tattccatta	tctacaactc	caggaatatt			660
attaatcaaa		gcttccat	ttt	cagtgggggt	tgaaatatcc	atattagaga	tatctaaaat		720
gtgattac	ct	tgggtctgtta	taaatccagt	tctatatgtg	ggtattccac	cgatcgagat			780
tat	tttttctt	gcaacaaggc	tcctactttc	aggtatcacc	tctataggca	gtggaaaagc			840
tcccaaaa	aga	ttaaccatct	ttgactgac	aactatacat	ataaactcgt	tagaggcaga			900
agcaactat	c	ttttctctag	tatgtgcgcc	accaccac	ct	ttaataagac	aattttcagg		960
agacac	ctca	tctgcac	cat	ctatgtaata	agctatatca	actacatcat	taaggctaaa		1020
gac	ctctatc	ccat	ttttcat	ttaataat	ttt	tgatgaagca	tctgaactag	aaacagctcc	1080
agcaa	atttg	tgcctatg	ct	cttttagttc	ttctataaaa	aaattaactg	ttgagccggt		1140
tccaata	cct	aaaatcatct	caggatgaag	attat	tttttg	atatattcta	tagcttg	ttt	1200
agcaac	at	tttgagc	cactcataga	gttataatac	aagaaaatat	aggtagttaa			1260
ttat	tttgag	actaaaaatt	aaaaaaacag	gttcttttaa	gaattcccag	aagtac	ctaa		1320
ag	cttattcc	taatgcatct	gtttatgacg	tagcattaaa	atcacctata	acatttgctc			1380
taaat	at	ttc	aaagctg	gggaataaag	ttttcctaaa	aagagaggat	ctgcaac	cta	1440
tat	ttttcttt	taaaaacaga	ggagcgtata	acaagattgt	aaatttatcc	gatgccgaaa			1500
agaagag	ggg	ggttattgct	gcatcagcag	gaaatcatgc	tcaaggggta	gccagtgc	at		1560
gtaagaa	att	aaaaattaat	tgcttgatag	ttatgccaat	aacaactcca	gaaataaaaa			1620
taaaag	atgt	aaaaagattt	ggagccaaaa	tactccaaca	tgggggacaac	gtagatgcag			1680
cattaaa	aga	ggcactg	ttt	attgcaaaga	aaaaaaaaatt	gtcttttg	tt	catccttttg	1740
acgacc	ctct	aacaattgct	ggccaaggga	ctataggaca	agaaattctt	gaagataaaa			1800
ataat	tttga	tgttgtcttt	gttccggtgg	gaggaggagg	tattctagct	ggtgtatctg			1860

cctggatagc acagaataat aagaaaataa aaattggttg tggtgaggtt gaggattccg	1920
cttgtcttgc tgaggccgta aaagctaata aaagagttat tttaaaagaa gtgggcctct	1980
ttgctgatgg ggtggcagta tcaaggggtg gaaaaaataa ttttgatgtt attaaagagt	2040
gcgtagatga agtcattaca gttagcgttg atgaggtctg caccgctgta aaagatatct	2100
ttgaagatac aagggttcta tcagaacctg ctggggcatt agcacttgca gggttaaaag	2160
cctacgcaag gaaagttaaa aataaaaaac ttattgctat aagttctggc gctaataata	2220
atttccaaag acttaatttt attggtgagc gatcagagat tggtgaaaat agagaaaaaa	2280
tattaagtat caaatccca gagataacctg gaagttttct taagctttca aggatgtttg	2340
gcagctctca agttacagag tttaactaca ggaaatctag cttaaagcga gcatatgttt	2400
tagttggtgt tagaactaaa actgaaaaat catttgaaat cttaaagtcc aaattaataa	2460
aagcaggctt cacctttagc gactttactc gaaatgaaat atccaatgat catctgaggc	2520
atatgggttg tggcagaaat agtgactcag gctctcataa caatgaaaga atatttaggg	2580
gagagtttcc tgagaagccg ggcgcgctgt taaattttct agagaaattt ggaaataaat	2640
ggmatatttc cttatttcat tacaggaacc taggttcagc ttttggaag atattaattg	2700
gcatcgagag taaggataaa gacaagctaa taaatcattt aaataagtca ggnactattt	2760
ttacagaaga aacctctaac aaggcataca aagatttttt aaaatgaaag gttaataactt	2820
taatctaaat ttaattgaaa aaagctcatc gctaggggtt tcccacggct ctttgaacaa	2880
ctcggattga gatctatcat cctcctcgtc gtaaattctc ccaccttag aatagaccaa	2940
aaatagatat gacaaaggag cgagctcata tttatatcta atttggaacg aagctacgcc	3000
agtattaaat tcattaacta tattatttcc tttataaagg tatccattag catctgaaaa	3060
aatactaata ggattttttg ctttcaaagc aacaaattga ctcttaagtc tgatctcatg	3120
tttattattt ttaaaccaat ttagatcaaa agataaggta tcttgtcttg aatcatatga	3180
ggcaagatta ttattatcct gccatatcag ccattcattt tcctttctta ttctatattg	3240
cgcattgatt cttaaagtat catttggaat tattgaacct gctatcttgt aaaattttct	3300

accataccca	ttcgaatccc	accgattatc	tttctctccc	ttaaagaagc	taactctcca	3360
gtcatatgtc	cagaatgaat	agttctttgc	ctcaaagtct	gctgtaatac	ctattcttct	3420
ttttgatttg	ataaaggggt	aggcttcatt	ttttcttggtg	atagttgtat	ttttcccaga	3480
agatctaaag	ttaaaatcta	attgaaatth	agagttgtcc	ttaaaactaa	aagaatthth	3540
ttgatcgatg	cctattggat	ttgaattacc	agctgtgtca	gcatcataat	ttagatcaat	3600
tccataatct	atttgththh	atatcgagct	attatcaaat	tcattththh	ttcgattthc	3660
tccaatacca	gcatgaatcc	agtctcttct	ttgcagataa	ccaaagtcac	ttaactcgaa	3720
gtcgtcttca	aaataaagaa	ggcttccact	tatgttagat	agthththth	gaagatatgt	3780
aaactgagtc	ctatacccaa	gcccattthh	accatctthh	tctgaggcta	acagatctga	3840
atatgtaatt	aathththth	aacgaatatt	gatgtaatca	ataacattga	ctgttgatga	3900
ttcgctgtgc	atttcattct	caacattcgt	taccatgaag	ccaagcgthh	taththcaa	3960
ctttgttcga	gatcgaagag	cataatagtc	tcttccaact	gaaaaggctt	catcagcctc	4020
acttgctaca	aatactccaa	attcattatt	attactthth	tgagtaagtc	ttaatgcaaa	4080
atcaatatca	gaatagthth	ththtagctgc	ctcgcaacct	tcttcattac	ththctctga	4140
gcaattatag	ctggggggcag	ctccaatcct	ccgcgtatth	ataaccgagt	atctatcata	4200
attactaata	tcaaatagtg	attggththc	attgaaaaat	gctctththh	ctgagtaaaa	4260
agththctga	gcagaaaagt	taataaccac	atcatcactc	tcagcttgctc	cgaaatctgg	4320
attaatagct	aaathththh	gacgacctth	tccagtgtca	taaaagatth	cagccccaat	4380
atctgaccct	tcttggtthg	taactgaatt	ththatttgaa	gatatatatg	gaaaaaaagt	4440
aagctthtgat	thtgatatagt	thtgththth	taagctatct	aactcttgaa	agtagtcatt	4500
tctactagcc	attgtthcgg	cactgtctaac	ccatgactca	ththtcggcac	tataacgtaa	4560
tgcggtgtaa	ttaaththth	thtatatcac	atcaggctgt	thcattaacg	ttacatccca	4620
aggaataaaa	aactcagaga	cccaataccc	atcaaathth	tgtgtththg	caatccaatc	4680
tccatcccag	tctgtcttaa	agtctcctgc	thgcgtthth	atggcatcga	aaagcgagth	4740

cccaagattt	atagcaagaa	tgaaagcttt	gttaccatca	ccatcaaagt	ctatatattat	4800
agagttttta	tcgctaagtg	aatttatattg	atctctaagc	gtcttcctgg	agaacataga	4860
atcattactt	tgaaaattct	taaatccaac	atatatacca	tccttatattg	agaaaattaa	4920
agccgttgta	agaagttcat	ttttcttaag	agtaaaagga	gatgtttcat	aaaaatctgt	4980
aatttcaa	gcattattcc	actcaggctc	atcaagagag	ccatcaataa	caattgagtt	5040
tgaccaaatc	agcacagagg	taagtaaaag	tgataatgag	gctaaagtta	ttttcataga	5100
tagattta	tcagagtatt	ttaatctatg	aagagtagga	aatcatccga	tcattctcag	5160
aaaaaacata	atcatattta	aggtcatgtt	tttctccaaa	atgcatatta	caaattctggt	5220
attgataaca	cagtccaacc	aataaaggctc	ttgataacctc	ctcgtttata	gagccaataa	5280
ttctatcgaa	atatccagag	ccatagccaa	gcctgtatcc	atttaaataca	actcctgtca	5340
taggaataaa	cattaaatca	atttcattta	tgttgacata	atcttcactt	ttaacctctt	5400
tgatcccaaa	ttgatttata	aagaagttag	gctgctcatc	cagcaaatta	aaagccatca	5460
tttcgtcatc	aattactttt	ggaatataaa	tattcttttt	aagtttagta	aaagcttgaa	5520
ttaacaaatt	tgtattgact	tcatttcgaa	aaggaaaata	taaggcaatg	ttttgcattt	5580
catgagtatt	gattttttct	aatacatttt	cttggattaa	aaaacttata	ttgctcttag	5640
ataaatcaga	aatagactgc	ccctgctcaa	agagtgattt	tcgtatttta	tttttcacca	5700
taaattgtgc	cgaggctaac	aaaccaaatac	accgccatgt	gctataacctg	aaccgatggg	5760
tcagggtggtg	aatttcataa	caaatacaggc	ttccctgtaa	aggtactgcg	caacaatgct	5820
agaggaatat	cacctattta	attgtatcgg	ttcaaattta	ttaacacatt	agcgtatgaa	5880
ccagaatgca	gttaattata	aaatatatag	gtattaagta	aaagttaatt	tttagagagc	5940
agactctatt	ttttgtatta	gcttttcgat	atctttatta	tttaaagcag	agtcagtatt	6000
ctcgggttggc	ttggataaaa	gcttactagc	catagttaga	ccagctatca	ccaaggcatt	6060
attcttatca	ctgatgccat	caagctcatc	atttaataat	tttgcagctc	taatgagttg	6120
gtcttttctct	tctggcggac	aggctaacgt	cagatctcgg	ccaaaaattc	ttaaagatag	6180

cgtttccatt tctgacatta ttttttgagt ttagttat	ctttatttag tatgagctta	6240
tcttttttcc agtctctttc cctttgcaca taagaatcta	ttattccttt ttgctcatca	6300
aatcttttta tcaaaagatc taccttatcc tctagttcaa	ataatgatga ttttttttct	6360
tctgacattg tataagttta attgtactag aatgaattgg	aaagtttctt taataaaaga	6420
gtaaaataag gcatggaaaa aataatTTTT	aaaaacagaa gagattcact tataaaacac	6480
ctacctaaga attcagcctt aattgtacct ggtgcagatt	tgcaatatag aaatgctgac	6540
tcattcttata atttaagaca ggaaagtagc ttctactatc	tgtctggctt ttgtgagccc	6600
tcttctctaa tggtttttagt taataatgga aaaagcattg	attcaataat ttttgttcct	6660
gaaaaagata aacttaaaga aatctgggat ggttatcgag	ctggccctga gggcgcaata	6720
aatgattttc tttttgatca agcttttgaa aataataaat	cagatgcttt aatgcctgaa	6780
atccttcaag ggctagaaaa agttttttat tcaataggga	agaaaaatgg ctttgatcag	6840
aaagtaattg actggacatg cgcagcaaat tctaaagata	ggcacagcaa atcaattgat	6900
attattgatg gctcttcgat ggtaggaaat ttaaggctta	tcaaagataa gcatgaaatt	6960
gaatattatga agagagcttg tgaaatttca gctgaatcat	atattgaggt catgaaatct	7020
ataaagcctg gggacaatga gcaggaaata gaggcgctat	ttttatatga attcgccaaa	7080
aggggaggaa ggtttccagc ttatacacct atagttgctg	gaggtgaggg tgcttggtga	7140
ttgcattata ttgaaaatga taaagagtta gcttcatcag	atttaatttt ggtagacgca	7200
ggatgtgaat acaaaatgta tgcattctgat atcacaagaa	ccttcccagt aagtggaaaa	7260
ttttcagatg aacagctaca aatttataat attgtccaca	aagccaatct tgctgcaatc	7320
gatgctgtaa aaactggtaa tagcataatg gagcccaaaa	tggtttcaga aaaagtaatt	7380
actgaaggtc ttgtagagtt gggatttcta tctggcgatg	ttaatcagct tcataaaaaat	7440
ggtgcattca aggactttta tatgcataag gtgggacatt	ggcttggact tgatgttcat	7500
gatgttggtg actacatgga gggagatgag tttatgaagt	ttaagccagg gatgataacc	7560
acaatcgagc caggcatcta tatcagtagc gcaatggatg	tagatgacaa atggaaaggc	7620

atcggcataa	gaatagagga	cgacatcctt	gtaacagatt	caggcaatat	taatctaaca	7680
gagaaggtgc	catctaatacc	tcaagaaata	gaatcattga	tggcttagac	tatggagggt	7740
ccaattgtta	tttctggcgg	agggataata	ggtaattaca	tttctcttag	gcttgaaaaa	7800
aataatatca	aaaccgttat	tgtcgaaaaa	gctagtagtt	tcaaagccct	agataagggt	7860
ataagaacag	tcactctcaa	tgagcattct	atgcaaatgc	taaaaaatat	tgggtatttgc	7920
ccatcaattg	ctcaaataca	cagcatcgac	gtattagatg	gtgagggtac	aggcaaaatt	7980
caatttctag	caaaggacgt	aggcagcgaa	aacctttcat	atgtaaccta	tttcaatgaa	8040
ttacaaaaac	taatttctga	tccatgtaaa	gaaagaacct	tatttaataa	tgagattgat	8100
tcagttcaga	atcttaatac	agaatctgat	ccagagatca	tgcttaaaga	tggcatgacc	8160
ataaaaaacga	atctaattgc	tggatgtgat	ggaagaaatt	caaatattgc	aaaaattgct	8220
tcacttacaa	gcagcttcga	tgactactta	caaacagctt	taacttttgt	cgttgatatt	8280
gataatgatt	cacatggcaa	agctcaccaa	gttttttctg	aaaaaggaat	atttgcactt	8340
atgccactcc	cagaaggcaa	gggtgagatg	aataaatgca	cagtggtttg	gtcaataaaa	8400
aatcaagttt	tgggagatga	gcctgtatct	gagtttgtaa	aaaataacat	ttcttttttt	8460
gaatcaaagc	ttaatgttag	tctcagggtt	aagtcagaaa	ttttaagttt	taaattatcg	8520
aaccatcatt	ttgaaaacta	tattagcgga	cctattgttc	ttcttggtga	tgctgctcac	8580
tcatttcacc	ccttagcagg	tcaaggtatt	aatctaggat	ttgcagatgc	agatactttt	8640
tgtgaagagg	taattagtct	ttataaaaaa	gggattgcct	ttaatgagaa	atcagtttta	8700
aaaagatatg	agattagaag	aaaaagtatg	aactttttta	tgttgaagtc	tatggacttt	8760
tttgtggatt	tatttggttc	agaaaattta	tatctaaggc	tgatcagaaa	tttaggtatt	8820
tcttcgttaa	ataaatcaaa	gtttgttaaa	gcattcttta	taagacatgc	ctctggaatg	8880
aataagtttt	aaatttgtat	taaacttttt	gaccttttagc	tctaagttct	ttaagaactt	8940
cactaatgcc	ttttttatca	atgattctca	taccttttgc	agatacttta	agatttacga	9000
acctgttctc	agattcaacc	caaaatttgt	gtgtgtgaag	attaggaaaa	aactttcttt	9060

atagttctga	gcttaaaatt	acaaaaggaa	cccgcgtgtc	agttcctttt	ggtaaaagga	10560
agattgtagg	agtggttata	aaaaatattc	aaaagccaga	tttcctaaaa	aaagcgggag	10620
ctataaaaaa	aattatcgct	gtacttgatg	aatatacctt	gttcgacaag	ccaatttttg	10680
attctatatt	gtggtcttct	gattactatc	atcatcctat	tggtgaggtt	tttaatacct	10740
ttatacccac	cgaattaaga	aaaattaata	ataaaaaaat	tgaagcttta	agagaatttt	10800
ctgaatattc	agtaaagag	gatgataaga	aattcgattt	aaccaaggat	caagaaaaag	10860
cagtcaaggc	gctttctaaa	tctaaaggat	tttcacccac	tttattatat	ggagttacag	10920
ggctctggaaa	aacagaagtt	tacttaagag	ttgccgaaac	ttttattaaa	aataataagt	10980
cagtattagt	tttagttcca	gaaataaatt	taacccccca	attgctctcg	cgatttgaga	11040
atagatttaa	tggtgaaatt	ggcatatatc	attctaagca	aacagcagct	aagagattaa	11100
agacttggct	aaaagctaaa	tttggttcta	taaaaataat	agtaggaact	cgatcttcgg	11160
ctttagtgcc	tttagataac	attggtttaa	taattatcga	tgaagagcat	gaccaatcat	11220
ttaggcagtc	agaagggttt	aaattctctg	ctagagactt	aagtataaaa	agggcacagc	11280
ttgcagatat	tccaattatt	ttgggatcag	caacccttc	gctgcaaact	ttaaaacttg	11340
taaaagaaaa	taaatttata	agagttgata	ttcctaatac	agttgatgga	aacaagcctc	11400
ctaaattaat	agccttagat	atcaataaca	gccctttaat	aggcggagtt	gctaaagaga	11460
tcaattgaagc	aatgcaatca	accatagaca	gaggagaaca	ggttctagtt	tttattaata	11520
gacgaggatt	cgctccactc	tatcaatgta	gtagttgtgg	gtgggtagca	gattgtaaat	11580
cttgtgatac	aaatttagtc	ttccaccagg	caagaaatag	attaatttgt	cataggtgtg	11640
aatctgccta	ctctgttaat	ttgtcttgtc	cggcatgcaa	gtctaataac	tttaatatgt	11700
atggagctgg	aacagagaga	gttgaagaag	ttcttaaaag	cagctttgta	aagactccaa	11760
taattagagt	tgatcatgac	tcaacaaaaa	aagtgggagc	tatggaggct	atagttaaaa	11820
aaattcattc	ctcagacgca	gcaattttag	ttggaactca	aatgcttgca	aaaggacatg	11880
attttcctaa	agtcacctta	agcgttattt	taaatgctga	taatggcctt	ataagcccag	11940

aaattaatgc attagagaaa atatctcaat tgcttattca ggtctctgga agagcaggaa 12000
 gaaataataa tcttgcaaaa gttattattc aaacaagata tcctgatgat ataaatctta 12060
 ataaaattaa gacaggagat tatatgaaat ttgcttctca atgcctaagt accaatgagc 12120
 aatgaactt acctccattt actactttat gtctgcttag gtgctcatca ccaactcaaa 12180
 agagtaatgt agatttccta gagaaagctg ttttaatttt atccaatagg actgatataa 12240
 atgttattgg tcccttgctt tcattagttt cgaagtcgaa aggaaattat aggcaccaag 12300
 tctatatcca tgcaccaaag aagacttttt taaataaggt attaaagttt ttgacaacag 12360
 agtttgaaaa atggccggaa tctaataagg ttaagtggtc tttcgacatt gatccaatag 12420
 acttaagcta aatattaatc ttaattaatt gtcttgggta tattgggtta ttgttttagtt 12480
 tattctctgt attaatttct tctacagtca ccccaaactt tatcgctatt tctgataaga 12540
 catccccctt ttgtattttg taagtcacaa agcctggatc aatactcata aaggatattg 12600
 ctttaggttt gtccttaaaa tagttatgta ttcctaggaa aacagatctt gcaatcattc 12660
 ttctccctgg cttgcccttt aatctttgtg cgtcttcagg gttgggtata aaccctgact 12720
 caaccaatac agaaggtata tcaatagact taagaactct gaaatcagcg tactcaacat 12780
 tctttttatg aatttttgtg aaagggctct ttttaagttg atccaatatc ttagttccta 12840
 aaattttact ttcagaaatt tttttcttat atatttctgg ataggtttgt cttgccgcat 12900
 cctcatcaaa atcaactggc tttagatttt ttatatcagc ctgtattctt tccctttggt 12960
 ttttagataa gttcctagca acagtacttg aagcttcatc tgaccatata aaaacagaag 13020
 caccctttac ggaagataat ctaaaccat cagcatggat tgaaacaaaa atatctgctc 13080
 catatttcct agcatcctgg tatctattat ttaaactctaa tgtctcatct ccatctctaa 13140
 tcatcaccgg tctgtacca taagtatctc ttaaggttct ttctaactcc ttcgcaataa 13200
 gtaacgtcac atctttttct aaaatattat ttggggcaac cgcaccaggg tatttaccac 13260
 cgtgaccgc atctatagca acaacaatat ctctaatact tttattgagg tttttatttc 13320
 ttttcacctt aagctctaata tttatatattt ctgtatttat agtctgagtt gggttttgcc 13380

gccacaaaat aaacttgcct cagatgcaag tattccggcg ccacccatgc ctggggagcc 14880
agcgcatatc aaaacttttc caaaattacc tttatggcta tcctttgctc tattggggcag 14940
aagatttttt aattcttgaa aggtaaaact ttgcaatata gacatgaaat tacttaagct 15000
aaattatggt gaatgctaaa aattcattct tcagcgaaat aaccaatcct atcggaatgg 15060
aatgatccca taaaaatttt ttcgtcatga ggcactgcaa tagtgggtag gccaaaaact 15120
gttctagaaa atgaaaactt atcttctctc ttaagtgttt tttgattaag tttatgaatg 15180
gagaagggtg gagaacaatt tgttttttct gcacaattac caaatcatt agcctgtaaa 15240
tctaaagaag ttaaccaaac cgaaccgtct tttaaaaaga tattatctgg actctgaata 15300
aaataactac ctgttttagt attttgatta atatcgtaaa cagataaatt gtcaccttga 15360
ttatagttaa cataaagaag tccagagccc tcacttagca aaattccatt aggcccactt 15420
ccatctgttc catcaacttt attaaaatta ttgtcactcc aaagtactac atgccctgaa 15480
ataactttta ataatgaatt cattaaccat tctttcatgc tgatatctct tttatacata 15540
tgagatgcat aaaaacttcc atctttttta agggcaacat cattgaaata atattgggtca 15600
gggaccctaa tacagccccg ccatatcata tcccatgaag attcattttt tataatttca 15660
aacatttcaa tcgactcaaa tggcgaatga ttaattacag ctagttgata gaagccttca 15720
tcattttcaa caagatctat tccgtgggga ttgaatatgt ctagctcgcc tcttatgcat 15780
gaagcgtcac cccaagagct ttctccaaaa gtaatttttg gtaccttttt ttcataggta 15840
tttaaatacca ttaaggcaaa gtaaccaggg gtatgttctg catatggacc gatccctcca 15900
aattcagaaa taaagaagaa tttattatca ggcgtaatca caatatcttc tgggtttgag 15960
aaattacaaa taaccttgat acggtcatcg gattcacact tactaatatc catttggggg 16020
cctatatagt cagtagatac gacagttaca gaaataaata aataaagaat agaaaccggg 16080
acggtaatct tataatagtg ctttataaaa atttctaaaa tctttgaagc atgatttgga 16140
agtgcgatca gccaaacccc ttttaaaaaa gatagagccc ccataacaac aaaaattact 16200
gaccatagcc tgtcagacca ttccggctga agaagaccag ttaaaaaaag aatcatccca 16260

aagaataaag ctaaatagcc tgagattttg actttcgcac caacaaatth cgaaaagaga 16320
 tatgtataaa gaggcctttat aagaactagt aaaccacagg ccaagaaaaa gaatgctaag 16380
 tagtaattca taagttaggt tttatataaa tgctccttaa taatactaac aagttctaag 16440
 ggcttgtcca atggaacatg gtgagcagct ccaggaaccc cctcaaaagt cataatgtca 16500
 ccatatgtat ttttaatat gtccaagata cttccggagg ttaataagct gtcttcaccg 16560
 tggatgaaca aagcagggca gccaaatgaa aatgtgtaac cgaataacct ttcaagactg 16620
 ctaaacaatga catcatcaaa tttccatctc caccagcct caatatthtt tactgagtgc 16680
 tcagcaatgt atcttaagta ccaatcattc gtacaatctt gcttaggcat taacctaac 16740
 cttttaataa tatctgtctt gtcttgatag tgcttgatca ttctgagagg agaagagtgt 16800
 tgattcgggt cataatccgg tggctctata aatgtatcaa taataataat attatthtatt 16860
 agatcctthc tttcagatgc aacgtaacca gcaacatgtc cgccgaggga gtgtccaaca 16920
 ataaaaatat ttgaaatat ttttttatcc ttttccttht caattacaga gacaatacat 16980
 tctccaaaat ctttaatgcc atatgaatct ctaaaagaag agtcacccat gccaggaaga 17040
 tctattgcaa ctatatthgc gcagtctcta aagtggggcg caataggatc ccaccattht 17100
 ttatgagcac ctgttccgtg aataagaatt attaaatctt tgctthcatc tttggagtth 17160
 cagctagaat aggatatatc cccatgagga ttcttgataa tctctgagct aggcttgtcc 17220
 tcaatggcat ctttgaacca ctgtggggca tgaataatgt cttgatttag attgttaggt 17280
 atttccataa acagtattct aagctataaa aaataaaaat atgaataaac ttaatttaac 17340
 gccagcagca actgtthttag tcctaaagga ttctcctgat gggatggaag ttttgatggt 17400
 aaaaagatca agtaggcctc ctttcggaga cttthttgtt ttcccgggcg gcaagattga 17460
 cgaaggatga ttcaataata agatagaaga tttttgtgag ggcgtgactg ataaagaggc 17520
 ctccataaat cttggattag attctggagg tctagcatat tgggttgcat gtattagaga 17580
 atgctthgag gaggttgga ttttactthc taaaaaaaag agtggggaag atcttgatct 17640
 agatggagtc gataaacata aatatcaaaa atatagagag atgttgthaa ataataaat 17700

tgatttatat	aaaatctgtt	tagaagaaaa	tttaattcta	atgcctcaac	aaatagcccc	17760
tttctcgcat	tggataaccc	ctgaaataga	aactaggaga	tttgatacac	gtttttttat	17820
tgcccacctc	ccaagcatc	agaccggaga	acatgatggt	agtgagctca	tagacagtgt	17880
ttggatttca	cctaaagaag	cgctcaaaaa	atctcgttcg	ggtgagatgc	ctatgattat	17940
gcctacaata	aaaaatttgg	aacaatgtgc	acaatttgat	tcgggctcta	agctttttaga	18000
aaatcagagg	aatctctcaa	atgaggatat	cccaccaatc	ctgccaaagt	tttttaaaga	18060
agatggtgag	tggagggggtc	tattgcctgg	agataaaggg	tatgaggatc	attaaataat	18120
atggacttaa	ttactaaaat	aacagctccc	aaccctgggtg	ttttcactgg	gggtgggact	18180
aatacttatt	tgattggcaa	agatgatata	acccttgctc	accctgggtc	aaatatatct	18240
gagcatctag	atgaaattat	caaagcaggg	gatgggaaaa	taaaaagaat	ctttgttact	18300
catacccata	cagatcattc	cccagccgca	ttgcctttat	caaaaactct	taatgttcca	18360
atgtacggaa	ggctagtaga	tggatgaatcc	tcattgggagg	atgaaacatt	tatcccagat	18420
tattattttaa	atgataaaga	tattattgag	acagacgaat	atacgttaga	agtaatacat	18480
actcctgggc	acgcatctaa	tcattttatgc	tttttaataa	aagatacgaa	atgccttcta	18540
acaggcgcgc	acattatgga	cgggtctacg	gttggttattg	ggccaccaga	tggcaatatg	18600
acaagctata	tcaattcatt	agaaaagtta	ctagattttg	atattgattg	ctttgcgcct	18660
gggcatggaa	attatattca	tgagcctgag	aaaaccattc	aatcaattat	taggcacaga	18720
ctaacaagag	aaagaaaagc	tcttagaaaag	ctaggagagg	caggaatctc	atcattagat	18780
aaacttacta	agcttggtta	tgatgatgtg	tcagagatgc	tccatcctat	agctaaatat	18840
agtctagaag	cacattttatt	aaagcttata	gatgaaaaga	aagttaaatt	agataaagat	18900
ctattcgaaa	taatttaatc	ctttttatatt	ttatgtaaga	ctttctcctc	aatagcttct	18960
atatcaatat	catcaattga	gtcttcgtta	ttatcaggta	tctttttttac	atcttttttca	19020
attttaaggt	cgataggaga	ggctccaaga	tcaaaagtta	gctcccttac	atttttttgaa	19080
atagtatcct	caacacagtc	atcttgatcg	tatgcctcac	gggtctctcc	tttatcatta	19140

aaatcagtc	accacttccc	aataccactc	ttgtctccag	aagctctctc	tcttacaagg	22080
agaaaatcat	atgcagctct	aaaccttgga	tgtcgaagag	ttttatatgg	ctggctacca	22140
attctgctat	gaagttttta	ttgaagtacc	caaatatcct	taatatagct	tgaaaatttt	22200
cttggtattg	ctgtgatttt	ttgttggtcg	cgaagtacac	catccataga	tcggaaaaat	22260
tttctcacat	taatttctcc	attcttagaa	cactttttca	ataacgaggg	ccataacaat	22320
gcagccataa	gaaagcctgg	tgtaattgac	tgttggtttt	taaccctatc	atcagtattt	22380
cttagagcat	gtgtcataac	attgctggca	aagtcattcc	tacttggaac	agaaaggata	22440
agatatttat	ttaaataaaa	tgagcagagt	ttttcaaaat	tcttttcacc	cattccattc	22500
aaaaatattt	tgcagaactc	atcaaacaat	ctagcgtttg	agatgcctga	taaaagatgg	22560
cctttatcat	agatggcatc	cttaaccaga	ttatctatct	taaaattgag	tttgttacta	22620
aatcttatag	ctcttaagct	tctcactgga	tcttcttcga	atcgtctttg	gggatctcca	22680
atagatacaa	taacccttct	gtgtatatgc	ttaagcccat	cgttatgac	ttctattttt	22740
tttgtaacag	gacagtagta	aagggcattc	acagtaaaat	ctcttctatg	acaatcttgt	22800
tcaagagtgc	ccaattatt	atctctaaga	atctttcctg	tcgaatcagt	aacaatattt	22860
tcaccatctt	cctgatcact	tcctgatctg	aaagttgcta	cttcaagcaa	ttcactccta	22920
ttaaaaacat	gaaccaattt	aaatcttttg	ccaattattc	ttgaagcttt	aaatgttttc	22980
ctgatctgct	caggcgtggc	attagtggct	atatcgaaat	cttttggctc	tagtcccgtt	23040
agtgcacgc	gaacgcaccc	accaaccaga	taggcctgaa	aattattttt	ttgaaggctc	23100
tgaacgacag	atatagcaaa	tttacttatc	ttattattat	ctatcaaata	taattatgaa	23160
tttttatgat	ttaagggcat	tagccattta	gctgcttttc	tttaatttcg	tcaagtgttt	23220
tgcaatgaat	gcaatgagtt	gcagttggtc	tagcttcaag	cctcttaatt	ccaatttcat	23280
caccacagga	ttcgcaccaa	ccataatcat	cttgcttaat	ttgttcaata	gatagaccta	23340
ttttgctgat	aagttttctc	tctctatctc	tggttcttaa	ttcaaaagca	aactcttctc	23400
cttgagaagc	cctgtctact	gggtctgcat	aggtttcacc	tttagctcta	agatgatcaa	23460

aagttttttg catttcatcc ttaagatggt ctttccaaag aagaagaacg gcaacaaaat 23520
gtttcttcat tgctgcactc atatatTTTT cacccttctt agatttataa ggtgcaattt 23580
tagatttatt attagctatt gttgcttttg cagacttctt ggctgcaact tttttaactg 23640
gagctttttt ggtaacagtt tttttaactg tagctttttt ggcaacagtt tttttaactg 23700
gagctttttt tactgtcttg gatttttttt cgaccatgta agattttata gaatttttagg 23760
gtggagaaaa tatcagatac tgacaaaatt agctagtcac tttttaattt atttaatact 23820
ttcaagtacc catccgagct aagtcttggc ccaaagtgtc caacaacctt ggaagatgca 23880
tagtttgcaa acttagcaca tgcttcaata ttattccctt gaaggtaggc atgcataaac 23940
gatccggcaa acatatcacc agctccattg gtatctattg gagttatttc ttctgcttga 24000
gcatgcttct caaccccttt atctataaca acacttccat cggcgctttt tgtaatagca 24060
gtcatatagg gcttttcttt ataaaagcta acagcatcat caaggctttc tttaccagaa 24120
aaagcaacag cttcatcatc attacagaag atcatatcta ttccatatga ctctattaaa 24180
tcaaattttt ctttaaaacc atgaacaata cctgcatcag aaagagacaa ggctttcttt 24240
acgtccttgt ctttaagggt ctctaagact gaaataacag cattaaggtt atcgtcactt 24300
gttaccatgt agccttcgat ataaaaaatt tttgaatttt ctacaacatc aaaatctata 24360
tctgatttac caagatacgc actaactcca agcatgctgc tcatagttct cttagcgtca 24420
ggagtaacta aaattaagca tttcccagtt ggttgatctg tattttcaga gctgacacca 24480
atatgtttga ctccagccga cctgagacta tcaagatagt ttcttccatc ttcatcatca 24540
gaaactctgc atacatgatg gcaattcgaa ccataatttg ctgcagcaac aagagaattg 24600
gttgcaagagc caccgcaatc agaaatcgat tcgggtccca tttcaataag tttgctaatt 24660
ataggtgcct gttcttcaga agatgaaaga gtcatagagt cggctacaag gcctacactt 24720
gataaaaatt catggctgac tttatataga gtatctacta aagcatttcc aagggcgcta 24780
atatcatatt tcatgtgtta ttcttttttg tattatttgt tttactgttt ctaacgttct 24840
attaatttct tgatccttat gcattgctga aatgaagcct gcttcatatt tagagggggc 24900

aaagtatatatt	ccacttctaa	tacatgaatt	taaaaaattt	gaaaatagta	catcatcagt	24960
ttttgcaaca	tcattaatat	tattagggag	ttcttctgaa	aaaaagaatc	cgaacattcc	25020
accaattctg	tttatagaaa	acggaatacc	tgactcaatc	attaagggtt	tcattcccatc	25080
caaaagcaca	gatgcatttc	tttctaactc	tttaaacgga	ttttctttta	tcaacaattg	25140
caacaaagca	gttcctccag	ccatagctag	tggattgcca	gacaaagttc	ctgcttgata	25200
gacaggacca	gaaggagcta	gatagttcat	aatttcttct	ttgcctccaa	aagctccaac	25260
aggaagaccg	ccacctatta	ctttcccca	agcagttaag	tcaggagtaa	tattataaat	25320
ttcttgagct	ccgcctagcg	aaactctaaa	gccgctcatg	acctcatcaa	atattaaaat	25380
agaattattg	gctgaggtgg	tttcctcaa	taactttaag	aaatcttcat	gacctggaac	25440
aaagcccata	ttccttgcta	ctggttcgac	tattacggcc	gctaagtc	cttttatctc	25500
attaaatatt	tctaaaaact	gttctttatt	attgtattcg	caactaaatg	tatatattgc	25560
caaactctgca	ggaacccctg	gagagtcagg	taagccaaag	gtggcaacc	cagaacccgc	25620
cttaattaaa	agagagtcaa	catgaccgtg	ataacagcca	tcaaatttaa	taattttatt	25680
ccttcctgta	aaacctctgg	ccaatcta	cgttgtcatg	gttgcttcag	tacctgaatt	25740
aaccattctt	atTTTTTcaa	ttgaaggaat	gcatttttta	attagcctgg	ccacatcaga	25800
ttcaagactt	gtcggggcgc	cataactagt	tccaagcgca	acttgatttt	taattgcacc	25860
tacaatgtct	gggtgtgaat	gaccattat	catcggaccc	caagacccaa	tataatcaat	25920
atattcatta	tgatcagcat	cataaaggta	tgcgccggat	gctctttcaa	aaaatatagg	25980
attgccattg	atatttttga	atgcccttac	tggtgaattt	accctccag	gcataagagt	26040
cttggcctct	ttgaataagg	ctatggattt	atcaatttta	ttatgtgtca	aagctaattc	26100
ttcttaattt	ttaatttcga	tatgatata	acttaattct	tattatgtat	ttaattgttt	26160
tgaaaattca	tccacagtat	tccagttgg	gaactcataa	gtatttgtaa	catctgtagg	26220
accctttgta	atccacataa	tcaatcta	catgtgttta	tctataaatt	tatattttgg	26280
ataatctatt	ttgcctgcaa	agactgctaa	tttcattggg	ttccaaggag	acagctctaa	26340

aaatttttgc atatacgggt ttgtttcagg tgtatttttt tcaggctttc tcgcaactac 26400
 attaactgaa aagaaggcat tttcttttgt ttcaagacaa gcaacatttt tttgaataaa 26460
 ttcataaagt tctggtttgt gtttgccata cctaattgctc gcaccaataa taattttatc 26520
 aaattgatat aaatctaact ctactgcttt tgcaatatgt attatttttg aagattcaga 26580
 cacatctaatt ttagagaaaa tttttttaca aatttccaga gtttgcccat cggttgtaga 26640
 gtagattagt agagttgatt tcataaatat attttatatt gttaattaaa aaatttcatg 26700
 gcacgcgcaa aatataaaca aaaatatagt atatttataaa aatgattaaa tattttattta 26760
 ttatcgcttt attcatcagc aactttgcat atgcaaatat tgatgcagct aggtgcgcag 26820
 ggattagttc agatcaagaa agacttgatt gctatgattt aattttttaa gctaattgatg 26880
 agctaccttt agatagtaat attaaaactc tcatcacacc agctatcaag gctgtgaccc 26940
 cagctgattc aataaaaatt gaaaataaag caacaaaaga aaaggatttc gggcttccaa 27000
 aaacaaaaat caaaaactct gcaaagaact caataaaaac ctcggtagta aggattaaaa 27060
 aaacaaaaag tggtaaatta atttttactt tagaaaatga gcaagaatgg actgctgaaa 27120
 cttcttatag agcaaggaat atgttttaac cagaaaccgc agtcatttta gaagaggccc 27180
 tggttagtgg tttttatatg attaataata gtaataaaca gaaaattaga ataaagaggt 27240
 tgaaatagca atgaccatac agagtatagc aacgaccgag ggtgccatga aaagaataag 27300
 atctgttctg gggcccaag atggctcaag cttcaggggt tatgttaccg gaggtgggtg 27360
 ctctgggttt caatatggct tcaagtttga taacgatatt gcattcgatg atgatgttat 27420
 aaattgtgga gatttctcgc ttttaataga ctctatgtcg taccctatc tttatggatc 27480
 aactctggat tttgttgaag atctctcagg ggctaaattc gttattaaaa acccaaatgc 27540
 caaaacaaca tgcgggtgtg gagagtcatt tacagtttag attttgtaat tgaacccaaa 27600
 agtcctttta cattcttgct tgatgtctga actatcagag cctcattact gactctttta 27660
 tatcccatcc aggcaaatgc catggactca atagcaaaaa catcatggcc taaatcactt 27720
 gagaggacga tatcattact agccatctca gaaattcttt ttaccagata cttgttatgg 27780

gcaccaccgc	cacaaatgac	aatgtcacaa	ttattatgac	catttttatg	gattgaattt	27840
attatagatt	ttgctgaaaa	ttcaaccagg	gtacatagaa	tatcttcagc	tttctttttt	27900
aagaatttct	tggataatat	ttttatatta	aagagctctt	tgcctgttga	ttttggacat	27960
tttcttttaa	aaaaattatt	ttgaagcaat	cttcctaact	caatgtgatc	tacttttctt	28020
ttggctgcaa	ttgcaccatt	tttatcaaaa	ggaatctgaa	gaaaatcact	acaatatgca	28080
tctaaaattg	cattaccagg	gcctacatca	gttccccata	tgtcatttct	attttttaca	28140
aatgagtaat	ttgaaatccc	ccctatatatt	aaaataattc	gtgggtttct	agctttataa	28200
aataattgat	tatggaattc	cggaacaaga	ggggcgcctt	cgccaccag	agcaatatgc	28260
atatttctaa	aatcacttac	caccagaagg	cccgtttctt	ttgcaacaat	atttggatca	28320
ccaatttgca	tagaaaatgg	aaatctctta	ttaatttcat	gcctaattgg	ttgtccagaa	28380
atagcaacac	actctataga	tgatttctta	attttactaa	atccaatcgc	ctcattaata	28440
gattttgaaa	ataagaatcc	tatctcttta	ttgatagtgc	ctaaatctga	caacgagctt	28500
tcgttatttt	caataagttt	ttttactttt	aatcttaggg	attttgggaa	tttaattgaa	28560
tggaaatact	caagatgtat	ttttgtacca	atgctaagaa	atgaaatatc	aatagcgtca	28620
tgactagttc	ctgtcattgc	tccaatatag	atctttttac	tcattactac	ttggagaaaag	28680
tttattgaac	tcctccattg	ataatttatt	attctttaat	agtgtatcaa	aacttgcctt	28740
ttaaattttt	ggtactgggt	cagctgatgg	taattttact	tttataggat	ctattctttt	28800
cttacctatc	ttaaactcat	aatgtaaatg	aggccctggt	gccagtcttg	aactcccaac	28860
gaaaccaata	gtatcgccct	gagaaacttt	tttgcctttt	cttataccct	tactaaattt	28920
ttctaaatga	caatatctag	tcgaatactc	atttgtatga	tttataacta	tctcattacc	28980
acatccattc	ctttgaccag	aaaaagaaac	aatcccatcg	cctgtagttc	tcaccggcga	29040
tcctcttttt	gctgcataat	cgactccatt	gtgagctctt	attgtatgaa	gaacaggatg	29100
cattctgttt	ggattaaaat	gagaactaat	atatgcaaaa	tctaaagggg	ctcttaaaaa	29160
ggccttttgc	atattatttc	cattttcatc	aaagtactgt	tttttactgg	cttcagtgaa	29220

aaacctatTT gcaaaatagg tattaccatt gttataaaac ttagcaatta cgatatcgcc 29280
 attttttact ttctctccat cactataagg agtgtcataa attacatgaa attcgtcacc 29340
 ctctctaata tcaaaaacga agtctacatc ccagccaaag atataagcaa agtccataat 29400
 cacactttcg ggaatatttg catctagagc tgcctcataa aatgaggatt gaataatccc 29460
 gctattgtaa gattcaatta actcaatact ttTgcttata ttcttaaata ctattttctgg 29520
 agtaaggcta attgaaatag aattttatttg atctttcatg atctcaattc tgggttagctc 29580
 ctcaccagaa tattcaaaga gcattttttc ccctgggtta atgttagcaa tgatattttt 29640
 ggagtctaata ctaaaaatTT tataagcagt atttaaaggc actgaaaagt tttcaaaaat 29700
 tattgaaaga ttttctccat cttgcacctc atgcatctga tatgttttcg ttTgctcgag 29760
 aggtatagag aatttttcag taatctcaat ttcttcaact ggcaaagatt catatgtttc 29820
 tatatcaata taaagcatta atattaagac aatagaaaca gcaaaagaaa caagtaccgc 29880
 tctttttTga acttttttaa aacctatcat tagtcctctt gaatatttaa gagttcagtg 29940
 ttgccaccaa atgctaccga gttctttTgag acaacctttt cgTtaagaaa ttgcagaagg 30000
 taattaggcc caccagcctt ataaccgctg ccggatagat ttTgtccacc aaaaggctga 30060
 gaaccaacaa cagctccac catatctcta ttTatgtaaa tatttccaac attacacttg 30120
 tcactaaata tTgcggccct tttttctact ctagtatgaa taccattTgt aaggccgaag 30180
 ccactatcat taatattttc aattagtTca tccatttcat ttgatttaaa tctaacgata 30240
 tgaagtattg gtccaaattg ctTgtctttt agatcagaaa tattatcaat ctcaattatt 30300
 gttggagcta caaaattttc atctacatta tcagcagacc taaaaataga ataattctta 30360
 tcttcaaaac ctttgacgta agcattgaga ctatcaagag atgttttgct aattattTga 30420
 ccgatatctg tTtcaagatt ttgagggttt ccaattttta gttctttcat accaccctta 30480
 atcattgata aaagatcTtc atatatTTcg tctTgtacac agagaactct caaggcagag 30540
 catcgtgtc cagagctatc aaatgctgaa cgaataatat catcggttTc ttTctcgaga 30600
 agTgcgcttg aatcaacaat cattgaattt atgcctcctg tctctgcaat caaaggaata 30660

ataatctcta	aactaacggc	cagtctgtct	tgctcctcag	catcgattgt	aatttctaca	32160
tcttttagact	ttgcatactc	tgtgagctga	ataagttttg	gaagtagatc	agacttttatg	32220
tcttttagct	ttttcatttc	atatctagga	gataaagctg	atatttttat	tgagacacca	32280
ttaattgtat	tttttgtcag	atztatcttg	ccgacttcat	ctatagcatt	cttataagac	32340
tgataataag	tgtcgggcttg	ttcagcgttt	cttgctgcct	ctccaagcat	gtcaaataag	32400
taaatttcat	tttctatat	cttaattttt	tttatgtcat	caaaatctct	gcccataaca	32460
aattcttgac	tgagaatatg	catggcacca	actacagcat	ttctaattgg	aaactcgcca	32520
gattttgaga	ttaaagaact	taatagggcg	cttgggtttt	tggtccactc	atcgggtgta	32580
gaaactacct	tgcttgcaag	aagaagaccc	catgttgatg	cattaacgaa	aacactatcc	32640
gctttgttta	aatgctctat	ccacgcaccc	tcagataact	tctcagatat	tattagatcc	32700
cttggttttt	tgtctgggat	tcttagtatt	gattcggcta	aacacatcag	agcaacacct	32760
tctttattat	ctagcccata	ttcgcttaga	aatgcatcta	gttttgttct	ctcgtcttta	32820
ttctctctgc	aagcatcaat	tattttattg	agcatttttt	gaaatcgaag	ggtcatttaa	32880
aaaatcagaa	ttactaataa	gatctgaaac	tatttcttgt	tcaggataaa	acttgttact	32940
tgtaaatgtc	atagctatta	ttttaatctt	atagtgccaa	gtatcaactt	actgattcta	33000
ttaatgactt	aaaaacatta	tgatagtcac	atgagtgatg	cacttaaatt	aattaaacga	33060
ggaaccgacg	agatcctcac	agagtctgat	ttaaaaaaga	aattagattc	tggaaaacag	33120
ctaataatta	aggcgggggt	tgatccaaca	gctccagacc	ttcatttggg	tcataaccgt	33180
ttattaaata	agctaagaca	ttttcaagat	cttgggcata	aagtaatttt	tctaattggg	33240
gattttactg	gtcaaattgg	agatccctcc	ggtaaaaata	aaactaggcc	aacacttact	33300
tcagaagaat	taattttcaa	tgcaaaaaca	tatgagaaac	aagtttttaa	aattcttaaa	33360
aaagaattaa	cagaagttaa	atttaattct	gagtgggtgc	acaagcttgg	tgagatgggt	33420
ttgattgggc	ttgcatcaaa	atataatgtt	gcaagaatgc	tggagcgtga	tgattttaat	33480
aaacgttata	gcgcaaatca	aagcatagct	attcatgaat	ttttataccc	ccttgttcaa	33540

ggatatgact ctgtagccct agaggctgat gtcgaatgcg gaggaacaga tcaaaaat 33600
 aatttgctag tagggagaga gctacaaaga tcctatggtc aagaacctca gggtgtttta 33660
 actgtaccca ttctagaagg cttggacgga ataaataaaa tgtccaaatc attaaataac 33720
 tttatagcaa tagatgaaga gcctaataatgat atgttcggta aaataatgtc tatttcagac 33780
 gagttaatgt ggagatgggt tgagttactc agctttacct cagagaaaga aataggagtt 33840
 cttagaaga aaatggaaga agggaccaac ccaagagata ttaagtttct ttagcagaa 33900
 gagttagtag atagattcca ttcagagggt gatggttcga aatgcaagga agcttttctt 33960
 caaagatttc aaaaagggtca aatgcctgat gacattcctt ccatgtcagt tgatgttggg 34020
 gctgagggca ttccattagt aaacctttta aagaattgtg agatgacatc aagtacatct 34080
 gaggcgatga gacttgtaa acaaggcgga gttaaaatcg actctgtaa aatagaagat 34140
 cctaaaatgc taatttcaa aggccaggag tctatttatc aggttggcaa aagaaaat 34200
 ttaaaaatta aaacataatg aaaaataaat tagtccaaat atttcttctt ctttttatag 34260
 ttgcgtgtaa tcaggatagc cctaataatta acaaattag taatatgcag tattttattg 34320
 ataatgaaat aagggagga atctattctg ttgaaccagg cttgcaatac tcaattatcc 34380
 aaaatggaga tcaaagttct gaatcaccat tgctgcagga tacaattaca gtcattttc 34440
 acgggacct cactgatggc tcagtttttt ggagttctgt tgaaatgggt gagcccttaa 34500
 cagtcgaact atcaggtcta atagttgggt gccaaaaaat aatctctatg atgaagaaag 34560
 gtgatgaatg gagagtttat atcgaccaa gtatggccta tggcgatgag ggccggcctg 34620
 ggataccttc aaactcaatt cttattttttg atattgagtt attagatata caaaaaact 34680
 aacccttata tataacggat agagcgtatc cataaacctc agcaacctgg ttaataat 34740
 tatttttagg tgttccggct ccatgaccag ctcttccttc aattctaatt aatattggat 34800
 tatcacacce ctgggactct tgaagttttg ctgcaaactt gaatgaatga gaggtacta 34860
 ctctatcatc tcttttagct gtagtaatca gagtagttgg atagcactca ccttcaacaa 34920
 tattatgcag aggcgaataa gctaataagt tttcaaactc atctttctta tctggagagc 34980

catagtcact ttcccaggcc cagcctatag taaatttatg aaatctaagc atgtccaaga 35040
cgccaacttg tggaattgcg actttaata aattaggatt ttgcaacatg gtcgcagcca 35100
ctagcaaacc tccattagag ccaccttgaa tagcggttga tgatggagaa ccaatttctt 35160
gtgctgtaaa aaatttagca gaatatgcaa agtcatcaaa aacattttgt ttattaaaaa 35220
gctgctgc atcatgccag ttatcaccat attcaccacc acctcttaga ttaactacag 35280
caactattcc cccttggttc atccaagtga ggtagctttt actaaagcca ggtagtcttg 35340
agatattaaa gccaccatat ccatataaaa gtattggagt attactgtct atttttaagg 35400
actttttgta actaagatgg atgggtatct gtgttccatc tttggatgga aagaatttaa 35460
agtcagatgt aaataaagtt gagtcgtgtc ctttaagatc ttctttccaa aaaagttctt 35520
gtgacatata agttagatta attttgtata tttctcgagg agttacaaaa tttgtaaatg 35580
aaaaataaga tacctcatct tcaatttcac caccaaaacc acccattggt ctttttcttt 35640
ctgttgctag cttattttta tatgctcctt ttaaatacaa gaagtgaacc tcagtaaaag 35700
tatcaacgag ataagaaact acaatagaat tatttataaa gctaacacta ctaatagaat 35760
tagtactttc accaacgact tcattccaaa caaatgagcc attttttatt gtaagtata 35820
ctacttttcc gttcgcagca ttttcagttg agtaaaacca aaaagtatca ttcttgcttt 35880
ctaaaaagct ataagcgcct attagctcat ctattaaagg aataaaaggc tgatcagggc 35940
ctagctgaac atagagtcta tttctttcat ctgtgccttc accaatagat agaaacttaa 36000
ttttagaatc ttttacaaca cttattcccc aactccatct tggcttttca ggattctcat 36060
aaacaatcac atcctcatct tgagcagtag caattttatg aaacattaac tttggagctg 36120
tattaatatc ttttaaaagc tcttccgatg gctcgtcgta tttttggtaa taaaatccag 36180
aatcatcatt ttcccatgaa gcaccagaaa atttagccca ttcaatccta tcatcaagag 36240
tttttctga ttcaatatca agtactttcc aggttctcca atcagaccca ccatctgata 36300
ttgagaaagc tagaagagat gcatcattac taacgctggc acttgcaaga gagatagttc 36360
catcttctga gaactgattt ggatcaagta aaactctatc ttggcattct tcacagtcct 36420

taatcatcag	cttgctttgt	tgccatgacc	catcattgaa	ataataaaaa	gtttttttgt	36480
taacctgata	aggcatgctt	atcgaatctg	tatcccaaac	ctcatctaaa	ttcttagcaa	36540
tagatTTTTT	atatttattt	tggcctataa	atTTTTgtgt	gaaattattt	tgtctctcga	36600
cccagtctgt	tgagtcctca	cttgtaaaat	cctccatcca	tctataagca	tcttcgatta	36660
ggtatccatg	gacttcttca	ctaaaaggaa	ctttatttga	ttcaggggat	tcaaaattat	36720
ctgtttgatt	tgagcaacta	actagaatta	atagagctgc	catgtatatg	tatttattgc	36780
gcataaaaag	cataatactt	tacaattacc	ttatgaatca	attcaatcga	gcatggcaac	36840
ttttacgcaa	aaaattggaa	gttttttatt	gtacttgccg	cacccatcat	gatgcttgaa	36900
atagcaatgg	catcaatggg	aacccaata	caaaacgtta	ctcagcctga	ggatattcta	36960
gaatttttca	atgagaatat	agcctttcta	ggctctgtta	gtctcttagg	cgtagtttta	37020
agtatggctt	ttatgggagc	gctttttgtt	tcgtatgcat	ccatagaatc	agaaaatgag	37080
attgagcctc	taaatgcctt	atTTTTtaggc	ataagaaaat	tctttccact	tctgggggct	37140
tatcttatag	cttcagttgg	tgtctTTTTT	ggtatcttat	tattaatact	cccagctttt	37200
tatgtagcag	caaggctttg	catttttctt	gcatttatta	tgcttgaaga	caaaggagct	37260
atagaatctc	ttaaattatc	ttgggaaaag	acggatgagc	atggcaccac	tttgtttggg	37320
cttaccatta	ccttcttttc	tttaacaatg	atTTTTgcat	cagttgccca	atccattata	37380
agtccgggat	taatgcaatt	agttgttctt	gcaattattg	aatatgtaat	agtaattcca	37440
tggggctatg	tatatttttag	tttatacaag	tcattaaaaa	gatattaggc	agattaacta	37500
gaataaatta	atacgaagct	tttcgtttac	aaaaaaacac	accactctta	taatccttcc	37560
tccgaagatt	tgcactttgt	taattcatgc	aaacaacttt	gggtctgtag	ctcagcttgg	37620
ttagagcgca	cccctgataa	gggtgaggtc	ggtggttcga	gtccaccag	accaccatt	37680
ctttgaacat	taacttttat	ttcctttata	cttattctta	actaaataag	gatagtatta	37740
tggctaaagg	tttagacaag	caaaaaaacg	acaagaagaa	aggcaaaact	ctaaaagaaa	37800
aaagagcggc	aaaaaaagaa	aaaaagaaat	agttttttgt	agaagactat	tattatctag	37860

gacatccttta	tcttttagaa	aagtttttga	aatacctctt	gcacagaaat	cacaaaccat	39360
tcctttcaag	cttacgatcg	ctatatctac	atttgagagt	ccttcaacaa	atgcatcaaa	39420
cctctcaggg	tttacatcaa	gatcttttcc	atcaaccatg	gtttcatgta	aatgaccttc	39480
atgagaatgc	ccatgcatac	cactatggtc	attctccgcc	gcaataaggc	cgcttgaaag	39540
caaagtaatt	aataataatt	ttttcatatt	tttctccttt	taaaatctgt	acataaaatg	39600
taaatcccag	ctcgtatctg	cgtcatagcc	aaactcaaca	agggcattac	cataaaaaaa	39660
tttcaattct	ggatacgtag	atgattttatt	tgttacagta	tttttttttg	ttttaaccat	39720
aacccatgta	tgtaaatacat	tgtagtcagc	aacgtaagga	gcaaataccaa	gttgaatata	39780
ttctttcttta	tagtcttttg	taaatttact	gttaatatcc	tttaacccaa	agcccgcata	39840
ccacttccgg	gtttccagct	ctccatggat	accataaaaa	ttatttttta	tacttccagg	39900
cacaattcct	gattgaaaat	ataaattcct	ttgggatgta	aaagtatttt	ttctattcac	39960
caaataagta	aaacgaaaat	agttttctgt	gttcttataa	aattttatttt	tagccctttc	40020
taggccaatg	gaatattttgt	aggttggaga	gtagtggaaa	taaatagaat	cattaaaggt	40080
atctgattta	tacattatgg	ttgttccgcc	agagtaggat	atcgggtcttg	catcagcatt	40140
aaagaccatt	aaaaaaacta	aatgatttat	tttatatata	cccatggggg	gtatcttacc	40200
ataggattat	aaattaattt	caggcatctc	tccaagcgcc	catttttcaat	gaaatcacca	40260
tcagaattcc	aaaaataacc	atagctatag	cagatgcata	aaaaatacta	actagcccat	40320
atgagagcca	gtaaacagct	accaatgcta	taacagaaaa	tactataagc	ctgatggccg	40380
tgattataat	tggccatttc	attgcattgg	ctccctgaga	tgcaaaatac	agggaaagcc	40440
catatccttg	aaatacataa	catacaccaa	ggatttgaat	atactgcttt	gtaactaata	40500
aggctctctg	atccgatgta	aaaataactaa	tccaaagatt	tggtgttagt	gctaaagcta	40560
gaccgataac	agccgaaagc	aatccagctg	ttgttgcacc	aaccatgcct	attttttctg	40620
ctcgtcttat	atttttggca	cctatatattg	ttccaaccat	ggctgtcata	gcagtcccaa	40680
tgccgaacac	aattggggatt	aataagaact	ctacccttga	accaattcca	taaccggcaa	40740

tagcagacgt	tccaaattga	ccaatgagtc	ctgtcagtaa	aagaactggt	gcaactgtca	40800
ttattggcga	aagagatgcg	ggtaaggcaa	cagaaaaaat	atcttcgaat	aattctcttt	40860
caattgttag	tcttttgagc	ttaagcctta	caggcgatga	agggctcgaa	atttttatta	40920
aagttactaa	tgccataaaa	cctgaagtaa	ctatcattga	ccaagcagat	cccacaagcc	40980
caagtttagg	tagaccgaat	gatcctaata	taaatccagc	agagaaaaaa	acttgtatac	41040
ctgcacaaat	taccgttagg	actgcaggaa	attgcatatc	tcccatccct	cttagggcgg	41100
cggttaagct	acctgatagc	caaacaacta	ttgctcccaa	taaatagaca	aaacaataag	41160
ctaaagactc	ctctaataat	gcacctgttc	cacctaatat	tcttaataga	ggctctccaa	41220
aaataaaaaa	gataattaaa	aaagctaaag	ctccaaagca	agatatatac	aaagaatgcc	41280
ataataattt	ttcagcccta	ggcttgtctg	cggctccaag	acttcttgca	atagaagatg	41340
tgaccgctcc	tccaagagca	ccaaatgcca	tctgctgagt	cagcatgatt	gcaggaaatg	41400
ccaaagtaac	agcagctaata	ggagttatgc	caagttgact	aataaaccag	aattcagcaa	41460
gaacaaccac	agcattttatt	aaaaatgcaa	cagtattagg	tgctgacatt	ttaatcagca	41520
aaggaaatat	tggatctttt	aaaaattgct	ctgttctctt	gtccataacc	ttctcattta	41580
ttaatattta	tttggttaaa	gggtgaaaga	gaacacgttc	tcaggattca	cgaatcttta	41640
aataataata	acagctttta	gttaatttaa	tttttaaaaa	caaccaaatac	tattatctgc	41700
tagattttat	tttttaaaaca	tattaatatc	taatttaatt	gcaaaggggt	ggccatttcg	41760
gcctgagatc	cttggaaact	ctttgaacct	gatccataca	ataatggcgg	aggaattgca	41820
tgaaaatcat	taaaaatcat	ttatttttatt	caataatctt	actatatttt	tttagtctag	41880
agacgagctc	tcaagctata	gaagaagtta	ttatcaaagg	agactggaga	gaaactagtc	41940
tgtcagcaga	agactcaagc	attgcagtcc	ttgatagtaa	attaatagaa	tcccaggcct	42000
taaagcactt	tgaaaacctt	tcgtacctag	taccaaattt	aaattttgct	gcaagtgatt	42060
ctagagcaag	acatttccag	ataagaggaa	ttggagaaag	atctggctat	gaaagaactc	42120
caaactctgc	agtaggttta	ctgattgatg	atatagactt	ttctgggcaa	ggtggaattg	42180

ccacaacttt	tgatgttgat	caaattgaag	ttcacagagg	ccccaagga	gcaagaatag	42240
ggtctagcgc	aatggcgggg	ctcatatata	tttctacaaa	agaccaaca	gaaagctttg	42300
aaggggaaggg	tgaaatagtt	atgggggtcat	atgggaacttt	taataaccgga	attgctgttg	42360
gcggtcctgt	aaattttaat	aaagacctca	cttatagggt	agctataaag	aaagattatt	42420
ctgacggatt	taggaagaat	atTTTTtctca	acaagtctga	tacttctaag	aaagatgaga	42480
gcacatttag	attaaaagtg	aattgggtaa	ctgataacca	aaccacttat	aagtttctta	42540
tatctcaaat	agagttagat	gacctgcag	atatatggac	tatagatggg	agtcttaata	42600
cattatctga	taagcctgga	atgggactctc	aaaaaagtaa	tgcttatggg	gtaaaaatTT	42660
atcatcaatt	taaaaaatTT	gaatttcaaa	gcctatcaag	cttaacgaat	actgatgtca	42720
ttcttagtta	tgatgctgat	tggggcaacc	ctgagtcgca	ttcacccttt	atTTatgact	42780
atTTTTcaga	aactacaaga	aaaagagata	ctTTtagtca	agaatTTaga	cttgtatccc	42840
aatttgcaga	taaaaataca	gaaaaaagca	tcgaatgggt	cgTTggggct	agTTTTgtag	42900
atataaatga	aacaaatgct	aaaaaagata	ctggTattta	tggagatcca	tcagacccat	42960
atggTcctta	TTTTtagtaac	tcttcttctt	tgagtgactt	ttcttcttca	agctattctt	43020
tatttggaaa	tattgattat	ttaattaatg	aaacaataaa	aatttcaatt	ggtggaagat	43080
gggagaattt	taaatctaatt	tattttgatt	cttatgacga	atcattttca	ccatcaaata	43140
aaatgtctgg	tggtaagttg	tcactagtta	aaactcttaa	taataattct	aatatttatt	43200
ttaatatTgc	taagggctat	aaccaggggtg	gattttaattt	aggtcttgggt	cttgataaaa	43260
attcatcaaa	tagaaattta	tattatgac	cagaatTTTT	aactaattac	gaagttggaa	43320
ttaatagtaa	atTTTTccag	tcaaaaattaa	atcttggagc	agTcctgttt	tattctgac	43380
gaaaagatca	acaggtctta	atTTcaaccc	aggttgatcc	ttcagatccc	aacactTTTT	43440
tatacttaac	ccaaaatgct	gcagaaggaa	ttaataatgg	cttagagtta	aatatagatt	43500
atgcactgaa	taaatctcta	ggtatatTTg	ctaattTTgg	attgctaaat	acagaaataa	43560
aaaattggat	ttcaagacca	gatatagaag	gtagagaaca	ggcgcatgct	ccaaaaaata	43620

gtttttcaat aggcataaat tggaagccaa caaaccaatc ttatttatca ttgaatgttg 43680
 ttggtaaaag tgagttctat tactctgatt ccataacaa tacttctgag tcatacaact 43740
 taacaaacat taattatgga tatgaacatg gacaatggac ttattcatta tgggcaagaa 43800
 atatttttga taagtactat tcagtaagag gtttttactt tggtaacgag gcacctgatt 43860
 ttatagatac gctctacaga agacatggag atccaaggca tataggagtc atgggtccaat 43920
 atgattttcta actttataag tgaatttttg atggaaatag cagctgttgt atttgctatt 43980
 atttatttat tacttgctgt aaaacaagac gttaaagtgtt ggtttgcggc tataattagc 44040
 tcaatattat atttctttat tatgtatgac gcaggcttat atatggaagc ttacctgcaa 44100
 attttttata tcatgatggc cttttatggg cttcaacagt ggagaagtgt cgatactgat 44160
 gctcctcaat ttattgttag aacatgggat aaaaggatgc atattaagat cattacatta 44220
 atagttatca tgactttaat ctctggattt ttattagaaa aatataccaa tgctatcttg 44280
 cctttcatag atggactaac aacatgggga gcaatagttg cgacatatat ggttgccaaa 44340
 agacttcttg aaaattggat atattggttt gttatagatt ttatttctat ctttttattc 44400
 atgtccagag gacttttatt aacctcagga ttattcttta tttatcttgt aataatatat 44460
 tttggttaca tgtcatgggt aaaaataaga gacgatatta gtgcagaatc ttcataacaa 44520
 tctagatata aaaaatcata atctagagat tattaaaaca attaaatcag gccctgtatc 44580
 tgaaatatct atttgtaatt ttgataacat caaagcaatc ttaagagttg atcatccatg 44640
 cgcacacaaa ataaatgtag atcgagaaaa tgaaattttt atactcagcc aactaaaaat 44700
 tttagatttt agtccagagg ttttatttag tgatttgtct tatggaattt tagtatggag 44760
 atatatcgag ggtattgaat tttcacttgg taaggattcc aatgaagttt ttttaaaaac 44820
 acttgaacc gaattaaaaa aaattcatga tattgatctc cccaaaagta agaaaaaata 44880
 ttttagcaat gacataaatt tttacagaaa tttactaaag gaggttcctg aaaatataat 44940
 tcttcataga ggatttgatt tatacgacaa actcaataac tctgataatt atgttctctc 45000
 tcataatgac ttgaataaga caaaccttct ttggagggat aggttatttt ttttagactg 45060

ggagtactcg agttttaaca atcctttttt tgatattgcc tcgttatcga atgcttataa	45120
cttatcaaaa gttgataggg caatttttatg gaaagcttat acaaataatg aatattcagt	45180
attaaatgat acaaatctta gagaatggat gcatttttgt cattatttag agtacatgtg	45240
gagtatttcg ctcatacaaa atggaaaaat tgatcagaac accctaaatt taaaaaaatt	45300
agagaaaaaa ttaaaaaata ttattttaac aacaagtatt gcgcattctt tagttgttgt	45360
tttactatta tgtaagaaat ttacatttaa atggggggtt gatatgaaaa ttttatgtgt	45420
cttatatgat gatccaaaaa caggatatgcc agaaaggat gcaagagatg atttaccaaa	45480
gtagataag tatcctgatg gaatgacact tccatcccca aaatctatag attttactcc	45540
tggtgagtta cttggttgtg tatctggaga actagggctt cgaaagtttc ttgaagatgc	45600
tggccataca ctagttgtta cttctgataa ggatggagat ggatgtgagg ctgataaaga	45660
attagtagat gctgatattg ttatatcaca acctttcttc ccatattatt taacaagaga	45720
caagatgaaa acagcgccta atttaaaaaat ggcaattaca gctggcattg ggtctgatca	45780
tgttgatctt caggcagcca tggataatag cgtggatggt gttgaagtta cttactgtaa	45840
ttctcgttca gtcgccgaac acattgtgat gatgatccta tcgatggttc gtgattatca	45900
taccaacat agaattgtaa aagagggagg atggaatata gctgatgctg tacaaaggtc	45960
atatgatgta gagggcatgc atgttggtac tgtagctgca ggtagaattg gtattgatat	46020
gctaagaaaa atgaaaccct ttgatgtgca tttgcattac tttgatattc ataaactctc	46080
tgatgaaata gaagcagaac taaacctcac ctatcatgat tctgtagagt cattagttgc	46140
tgtgtgtgat gtagttaata ttagttgccc attgcatcct aaaactgagc acttattcga	46200
tgatgaaatg attagtaaaa tgaaaagagg tgcatatatc atcaatactg ctctgggcaa	46260
gatttgtgat aaagatgcta ttgcaagagg cttagagtca ggccagctaa gtggttatgc	46320
tggtgatggt tggttcccac aaccagctcc aaatgatcac gtatggagaa caatgcctaa	46380
ccacggcatg actcctcata cttcaggtac ttcgctatct gctcaaacaa gatatgccgc	46440
tgggggttaga gaaattctag aatgttatTT tgcaggcgaa ccaattagag acccatattt	46500

gattgttcaa aatggtgatc ttgcaggtat ggggtgcgcac tcgtacacaa aggggtacagc 46560
cacagatggc tcagaagagg ccgctaagta taaaaaataa gttttagaac ttacttagcc 46620
tttccttaat aatagagtct gcctctgcca taatgctatg cattaattct tccactgtcg 46680
ggatgtcatt aaccagtcca gcaaccattc cacatgacca ggctccaacc tccatagttc 46740
cttcatgcat aatTTTTTgga tagactcctg caacctcatc cacaatatca gcaaaagtta 46800
attcatcacc aagagctttt tctttttcaa tcaatctctc aacagcttca ttattaagaa 46860
ccctttctgt atttgTTaat gatctcatga tgagtctagt atctaactca gaagcattta 46920
cgatagcctc tttcacattt tgatgaacag gtgcatcttg agtagcaata aacctagttc 46980
ccatattcat tccctcagca cctagtgaca tggcagcaac caaacttctt ccatctgcca 47040
ttccacccga ggcaacaaat ggtatttcaa gctcatctgc tgctctaggt aagagtatga 47100
aattaggaat gtcactttct cctgggtgtc cgccacactc aaaaccatct acagaaaccg 47160
catcgcaacc tattgcttgt gcttttaatg agtgccttac agaggtgcat ttatgaataa 47220
cttttattcc tgcttctttt aaagctggga ggtattctgc aggatttctt ccagcagtct 47280
ctacaacagg aacccttgca tcaataatta cttttatcaa accaggatag tctgggggtg 47340
ttagtgatgg taaaaatgtt aaattaacag caaatggctt attggtcatt tctttgcatc 47400
tggcaatttc atttgctaatt ttctcaggcg taccctgtgt tagaccagta attgttccaa 47460
gtccacccgc atttgatact gccgctgcaa gctcggcaaa accaacaatgg tgcattccac 47520
cttgaatgat tggatgttct ataccaaata attcagttat ttagtTTTTc ataatactct 47580
cctattttatc ttccattggg gttaaaaaat cttcatattg agtttgaagt ctttgcatgc 47640
cacttatcca tctatctctg tcattacctt ttcttttaac atattctaga acctcactat 47700
gaggggttac taaaaacctc tcctcctcta tggcgTTaag aacatccttt gcaacaatat 47760
cagcttccat cattccatca actcctgcta caccaggacc atttgcagtc atagcagttc 47820
taactgcttg agggcataaa caggaaacgc cgataccttt atttccatat gtaattttta 47880
tccattcagc aaaactcaca gctgcagctt ttgtaactgc atatccagcc gcacctagtt 47940

gagttaaaag accggctgct gaggatgtat tcataagata gccttcacct tgttctatca 48000
 tttgaggaag cacatgtttt gcagcatgaa tatgagactg aacatttaca ccccatatca 48060
 tgtcccaatc cgaagtatct gcttcaaaaa atcctggctt tccaccgata cctgcatttg 48120
 aacaaaatat atcaatacca ccagaaaatt cattagcctt ttgtataaca tttataatgt 48180
 cgttttcttt ggaaacatca gcacttactg caagcccatt aacactcttg gctgtttcct 48240
 ccgccccatt taaattcata tcaacacata ctatagaact tgccccagat gcatagaact 48300
 cttcacataa agccttacca attccactgg cagcccctgt aaccacaact cttttattat 48360
 taattttcat aaattgacca cttttttaat tttttattta actttattgc ttagatagtt 48420
 tattacaaaa caacaatact taatattgca aaatcgcata cccatcttta taaatttatt 48480
 gctattatta gtgctagaaa atgaaatact gagatttaat atgaagaatg ttgttggtat 48540
 tggctcgtcc ggggcaatag gaaaagcctt tattgatagc tatatcaaag atgatgatgt 48600
 tgaaaatata ttttcatttt caagaacagg cctttccatt gaggataaaa aactccatag 48660
 tttttttatt gatattgagg atgaaactag tatttgtgat gccgcagaga agatagacaa 48720
 gtcctcaata gatgaaatta tcgtcgcaag tggaatactt cataataaag attttgggcc 48780
 agaaaaaagt attagagatt taaatgcaga taacctttta aaggtcatta aggttaatac 48840
 tatcggccca acaattgttg gaaagtattt cattccattg ctaaataaaa aagaaaaaag 48900
 cgtcttagca tttttaagtg caagagtcgg cagcatttct gataataaaa caggtggttg 48960
 gtatgcctat agagcgagta aaactgcact taatcaaatc attaaaagtt ttagtattga 49020
 attacgaaga accaatccaa atgccattat ttttggtcta cagccaggaa cagtagatag 49080
 tgaattaagc gaacctttta aaagaaatgt aaaagaaggt aattttattta ctccagaata 49140
 tagtgtattg cagctaaaaa atattattga tacagcaagt ccatctgatt caggcaaact 49200
 aatttcttgg gatggggaag agattcagcc atagttgggt atgaatatat tttcatatta 49260
 gaataaattt ttaggggaaa aaatgagtat taaatattat gactggtcta aatttcaggc 49320
 caacactagg ccaaataaag ttgccataag agagctagat aataacaaga tctatactta 49380

cggagaattg gataaaagat catcaaggct tgcatacat ctccaaagtt caggaataaa	49440
aaaaggagat cgtattgcga tactatcgct aaattgttca gaattttttg agctagaatt	49500
tgcttgcgga aagattgggg caatagagat accattaaat tggagattaa caaaacccga	49560
gctcagttat attcttaatg atagtgaacc aaaaactcta atttatgaca atcagtttga	49620
agaaatggtg aaagagctaa aagaagaatg taatatcttct gaaatcatag ctcttgatca	49680
at ttgacca gaaagtgatt atgaaaaagt tttagtaat gcttcaggca ttattatca	49740
ggaagaagtt gatctagaag ataacattat gattatgtat acctctggaa caaccggtca	49800
ccctaagggc gccatgatca cgcacaaaat gcagcttttt aatgttatta atttaggtat	49860
ttcagcagct gtttccctg aatcagtcca tttagttgtc ctccctttat ttcatacagg	49920
cggaatgaat tgttattcaa atccaattct tcatgcaggt ggcgagttaa tattacttaa	49980
agagtttgag cctgggaaag ttctatcaat tatcggcagc tctgactatg gagttactca	50040
tctgtttgca gttccagccc cttatcaatt tatgatgaat catccagatt ttgaatcaac	50100
aaatttatca ggagttaagt atgctggagt tgggggcgca ccttggtcag aggctat ttt	50160
gaagacttat ataagcaagg gtgtttcgat gcagcaagga tgggggtatga cagaaactag	50220
tccaggtgct actggtcttg aatcgtccga ggctgaaaga aaaataggat ctgctggaaa	50280
accgcttctt catactgagg tcaagggtgg tggagatgat gggaatgaac tgcctgctgg	50340
agaagtaggc gagatttata ttaaaggccc aaatattaca cctggctatt ggaagaaaga	50400
agaggctact agagattctt ttgaagacgg gtgggttaaaa acaggtgatg ctgcttactt	50460
tgatgacgag ggttttttat acatagttga tcgatggaag gatatgtata tctcaggtgg	50520
agaaaatggt tatccagctg aagttsaaaa tggtatctat cagttaccac aaatcgcaga	50580
agttggagtt attggtattg atagccctaa gtggggtgag actggtaaag cctttgttgc	50640
tttaaagccc gatcatgaat tgacggcaga agaagtcata gatcattggt taaaaaatct	50700
agcaaagtac aaaattccag agaaagttga gtttattgca gctcttccaa gaaatgctac	50760
aggtaaagtt taaaaagaa cattaagaga tatgtaatat ttttaagcca aaaaaaacc	50820

agcctaagct gggtttttta tttaactatt taaagttatc taatgttaag cattagaaga 50880
 ttctttaaca gcaacattcc atataattaa accaaataga atcttgtaa caaagtcagc 50940
 aaggttatag ataaggttta agttaagagc tgatccaccg tcacccatca ggtaacctgt 51000
 gaaataacct acaggataaa tcgccaacc aaagatgata atatacatca ttgtgttgta 51060
 agctgattgc acagcaggac ttgcagtatt acatgcagat tttccttctc cagcccataa 51120
 ttcataaatc atgtataccc aagctaaaca cccaataatg aatgcaggcc atgcagccat 51180
 gattcctgct tcacccatgt aaccaaacac aagcataaca agagaaccaa ctagtaatth 51240
 cttaaataat gatccagcaa cattagttgc agcagcaaga attaagtaga attcacatat 51300
 taatagagga actgttagta accaatcaat gtatctaat acagttggcg aatcaccagt 51360
 ttcaatccat acccctctca tgtacatgta atgccagaaa gcaataccag taacaagacc 51420
 agatacagtt aatgatgttt tccattttgc agaaactcta tctctttcaa caaagaaaaa 51480
 tacagtagat gctaataaag cagcagtaac taaccaaaaa gaaacaccag tgtaatcact 51540
 agcatcaagg tcaccaccac ctgcagcaaa thtaggaagt gcaataacac tacctaatat 51600
 cagtaataat ttcatatata actcctatat taatgacaaa gaagcataat tgcctcccc 51660
 ctaattaaat atgaaattta atagtaacca atattaatag aggttacaaa taaaagcat 51720
 ctattttttt aatagaaaat atgtgtaata aaacaccata ctttaaatta atatttattt 51780
 tgcataagta gacttgttta gtatcatatt taacatgaaa gttgcaattt atcctggttc 51840
 ctttgacccc atcacaaatg gtcacacgga tattattgat aggggttgcg gactctttga 51900
 caaggttggt gttgcaatag ctaagagtga atcgaaaaac cctcttttta gcctagagga 51960
 tagaattaat ttagcccaat ctatttttaa aggaaatgaa aaagtagagg ttgttggttt 52020
 tccaagaaag ttaacagttg atcttgcaaa agactatgga gcttggtgcaa ttataagagg 52080
 cctacgagca gtttctgatt ttgaatatga atttcagtta gcaacaatga ataggctcgt 52140
 ggctcccaat attgaaagca tttttttaac accaaaagaa agtctcattt atgtatcttc 52200
 tagcttaatt aaagaaatat cagacttaaa aggcgatata tcgaagtttg ttcatectat 52260

agtagagcag	gcacttcgag	cgagtagaca	cttagctctg	acaagcttca	caaaaaaaag	52320
tagctctttg	attgacgac	gtttttacaaa	tagtgccctt	gcattttattg	cagggctcgc	52380
cttctcttcc	atatacattt	aattttaaatt	taaaatatcc	tggactacca	tcagctgagt	52440
agaagtcttt	taatgtagtt	cctccaacct	ctattgcttg	ctctagtatc	ttttttccag	52500
cagctactaa	tctttttacaa	gcactctaaat	ctaactcatt	ggcattttttt	agcggatgaa	52560
tttttagctaa	gaaaaggctt	tcagacgcat	aaatattgcc	gataccgact	acattttttt	52620
gattcattaa	atagctttta	atatttgttt	tagagtgact	gcactcctgaa	aaaaaatcct	52680
ttgcattaaa	attttttagaa	agaggttcag	gcccagggtt	ttttattaac	ttatgtttat	52740
cgatatcaga	agtaagatgc	attgatccaa	accttctggg	atcattataa	ataattcttt	52800
cctcatcaaa	aattaattca	atatgatcat	gtttgataaa	gaagttttca	ttattttttg	52860
caattctaag	actcccagac	attcctaaat	gaagaataat	ttttttatca	ttagatagtt	52920
taaaaattat	atattttgct	cttcttttcta	agcttttcgac	aacttgatct	tttactgagg	52980
tttcaaagct	atccaccact	ttccatctaa	ggtttctggt	gtgtatcctt	gcttctttca	53040
gcaaagagcc	tttaaatttt	ttaattgctc	tcagggttgt	ttcaacctct	ggaagttcag	53100
gcatttttaa	ctgagtaaat	tattaatttt	tactatgtca	gctggagtaa	ttgttcccga	53160
tgcgagccct	agtcttagat	tggcaagtat	gtaatcatat	tttgcattag	caagattttt	53220
ttccgcgctg	tataagtttt	tttctgcctg	caagagatca	acaacgtttc	ttgttccaac	53280
tctgtagcca	acttgagtcg	cttccagggc	actagtggcc	gaaatcactg	cttgtttttg	53340
agcatttaca	tttgcaacta	atgttaaaac	atttgaaaac	tgggatctga	cttcttgaat	53400
aatccttctt	tctgtaaata	gagtattttc	atttgctctt	tcatactgtg	aatatgcttg	53460
cttccttctt	gagttaacgg	cgccgccttg	aaagagtggc	atacttagct	gaattgcata	53520
attccttctt	cctgttactg	atggaactgg	aataccttgg	ccattgatat	taaaaccttc	53580
atagttaaat	tggtttggtt	cagattctga	ctgacttcca	acaatgtcta	tcttaggtaa	53640
atgatttgaa	gctacacttc	ttgcactgct	tttcgctgct	ttctttctca	aatatgctgc	53700

ttttaactgg tagttat	ttt ccaatgctaa	ttcaacccat	gtctcttttg	aacttggtgt	53760
tggcaggtca ataagcaa	acatctccca	ttcatttaag	ctgaatattt	ctctaccaat	53820
cagagcattt aaagactct	cttgcagaata	aagtgatcct	tctgttctaa	ttcttgaggc	53880
tttacttaga tcaa	atgagcctc	ttgaactcca	gttatggctg	ataacccaac	53940
atcgaatctt tgttttgct	gatcaagttg	ttttttaata	gctttttctt	cagatattgc	54000
tgcatttaga ttatcaatag	ctctaagtac	gccaaaataa	agctcagcag	ttcttactaa	54060
aagattttgc tgctcaa	atgcaaagtctgc	ttcagcagca	tctgtaagag	atttagattg	54120
cctatatattga aaccatgtat	ctagtctaaa	gagcggctga	gtaaccctgg	cagatgtaga	54180
aaaagagtta tattgctgct	gcagttcttt	gttttgatag	tattcggtcc	agttagttga	54240
tccactcaag gtaatactcg	ggagaagcgc	agctcttcct	tgaaccttaa	gctctttatc	54300
tgctaaatat gaatattccg	ctgctttata	tgtaggggtca	ttctcaagcg	cttcattata	54360
gatatccaaa agactttcag	atgaaatatt	aaatgagata	aaaagtgcta	aaaatgattt	54420
tgtgtaaatt ttcataactt	attttaacct	attaataatg	tttgcagtgc	aaacatttat	54480
tttttttaat tttattgaat	ttattttata	tttcttagag	tagaatatct	ctaacaagtt	54540
caataatttt ttataataaa	aaggctttta	aaattggcta	aaaattcata	tgacgctcag	54600
gcaattgaag tcttatctgg	actagatcct	gtcaaaaaaa	gacctgggat	gtatacggat	54660
acatctaacc caaatcattt	aattcaagaa	gttcttgata	attcggttga	tgaagctctt	54720
tcagggttatt gttcaa	acataaaagtatct	gttctaaaag	atggctttat	taaggctctt	54780
gatgatggaa ggggtatgcc	aattgatgag	cacccggaac	ataaagtttc	aggtgttgag	54840
ctcatccttt gtaaacttca	tgctggagcg	aaattctctg	gagatgatta	taatttctct	54900
ggcggccttc atgggtgttg	gagtttctgtt	gtaaatgccc	tatctgatga	attagagggt	54960
agagtaaaaa gagattctaa	agaataccaa	atcactttta	ataatggaga	taagtcttct	55020
gaattaaagc caattgggga	agtggggctc	agaaattctg	ggacatcaat	taaattcaaa	55080
ccaaatccta catattttga	aactatagag	attcagataa	aacagcttaa	gcatttatta	55140

aaggccaaag	cagttctctg	tcctggatta	acgatatagagt	tcgtttaatga	aaaaaagact	55200
gatgataaac	aaaagtggta	ttttgaagat	gggctcaaaa	gctatttgat	tgattcttcc	55260
gagggagcag	acttggtttt	gctagattca	attgtatgct	ctaaaaaatc	tcaagctcaa	55320
gagcttgaat	ttgcaatcaa	ttggtcatta	agacccccaa	aaaataaact	cgatgaaacc	55380
tatgtgaatc	tcataccaac	tgctcagggt	ggctcacatt	taaatggctt	taaggctggg	55440
cttttagatt	cattaaaaga	attttgtgaa	tacagaaatc	tattgcctaa	aggttttaaaa	55500
attaatgcag	atgatgttct	taataatgca	atttttataa	tttcatctaa	gcttcagaat	55560
cctcaatttg	cagggcaaac	caaggaaaga	ctagattcaa	aagatcacat	gtcgttcgtc	55620
tcaagtacca	caaaagacat	tttaagtatt	tggtttaaca	ctcatacaga	agagggcgaa	55680
agaatagcag	aacttgcaat	tatgtctgct	cagacgagag	caaaagtttc	caatatagtt	55740
gaaagaaaga	aaacttttag	aggcccagcc	ttacctggaa	aactttcggg	ctgtaatagt	55800
caggacttaa	atgaaacaga	gcttttttta	gttgaggggg	actcagccgg	agggtcgcga	55860
aaacaagcaa	gagaaagatc	tttccaggca	atcatgcctt	tgagagggaa	gattttaaat	55920
acttgggact	tagaaagtgc	agaaataata	aaatctcagg	agataaaaaa	cctatcaact	55980
gcaattgggg	ttctgccagg	aaataatgac	ctttcatcac	taagatacgg	aaaaatttgt	56040
attcttgcag	atgctgattc	agatggtctg	catattgcaa	ctttactttg	tgcatgtgtt	56100
ctaaggcatt	ataaatcttt	agttcaagag	ggaaggatat	atatttcaat	gcctcctcta	56160
tatagaatcg	attctggtaa	agatgttcta	tatgcacttg	atgataaaca	gcgagatgaa	56220
atagttactg	aatttaaaaa	gaagaagggc	aagcctaaag	taaacattca	aaggttttaa	56280
ggacttggtg	aatgaatcc	acctcaacta	agagagactg	tgatggaccc	tgctactcgt	56340
cagcttgttc	agctttctat	cagctcaagc	gataatgcaa	attctatgat	ggacttactt	56400
ttgtccaaaa	agaacgcacc	agcaagaaaa	gaatggcttg	aaaagaaagg	gtctctagca	56460
aaaatataaa	tatgaaagaa	caaataacct	caattagcct	caagcaatat	gctgaagagt	56520
cttatcttaa	ctatgcaatg	tatgtcattt	tagatagagc	tttgcccta	attggagatg	56580

gccttaagcc	tgttcaaaga	agaatactct	atgcaatgtc	agagcttggg	cttgatgctg	56640
gctcaaagta	caaaaaatca	gcaagaactg	ttggagatgt	tataggaaaa	tttcatcccc	56700
atggagacgg	cgctgcatat	gaagctatgg	ttttaatggc	tcaaaatttc	tcattcaa	56760
acccttttgt	agatgggtcaa	ggtaactggg	gttctcagga	tgatccaaaa	tcttttgctg	56820
caatgaggta	tacagaatct	aagttaacta	aatttgcaaa	tcttttaatc	tctgaattga	56880
agtctggaac	agtcgattgg	cagcctaatt	ttgatggctc	tcttttagag	ccagtaattt	56940
ttccagccaa	actcccatct	attttattga	atggcacttc	tggaattgct	gtaggaatgg	57000
caacagatat	tccatctcat	aatattaatg	aaattattga	tgccacagta	catcttattg	57060
ataatccaaa	atcacagttg	gttgattttac	tcaagataat	taatgggtcct	gattttctcaa	57120
ataattcgcc	aataattgct	agcaaagatg	agctgaatga	aatttattcg	actggaaaag	57180
gcggttcaa	agctcaagcc	caatggggcg	aggataagaa	tcaaatcatt	atcaacgcat	57240
taccttatca	agcatctggg	tctaaaattt	tagagcaa	agctgatcaa	atgcttaaga	57300
aaaaaattcc	aatggtgggt	gatcttactg	atgaaggaga	ccacaaggag	ccagtaaggc	57360
ttgtcataac	tttaaaatcg	aacagagtaa	atgctgaaga	tgtaatgaat	cacctttttg	57420
catcaactga	tttacaaaaa	aattatagag	taaatatgaa	tttgatttca	ttgaaagggtg	57480
gaccaaaggt	tttctcccta	gttgatttat	tgaaagaatg	gctagtcttt	agaaaagaaa	57540
ctgtaataag	aaaactagaa	catagactcg	accaggtaaa	cgataggctg	catatccttg	57600
aggggttatt	aattgtttat	ttagatttag	ataaagtaat	aaagattatt	agagaatcag	57660
atgagccaaa	aaaagatatt	attacagctt	ttaaactttc	tgatatccag	gcaaatgcaa	57720
tccttgagat	taggttaagg	caactagcca	agctagaaca	aatcaagtta	gaacaagaaa	57780
gagatacttt	ggtagcagaa	caaacagata	ttgaaaaaat	actaagctct	aagacaagac	57840
ttaaaacatt	aattaaaaat	gagcttattg	aaataaaaaga	tgagtttgga	gaagtaagam	57900
aatctccaat	aagagaagca	acagaagcca	aagttttttc	tgaagaagaa	actcttgctc	57960
ctgagcctat	tacagtagtc	ttgtctgcag	ctggctggat	cagaagtgca	aaaggccatg	58020

agatagaccc	cagctcactt	tcctatagag	gagaagatgt	acttcaagat	tatggaagag	58080
gaaagagcaa	tcaagtttca	gtttttcttg	actcaaatgg	gaaggcttac	tcacttgcaa	58140
gtcactctct	tccatctgct	agaggaatgg	gtgatcctat	tacaggaagg	gtatctgcag	58200
attctggagt	aaagttcatt	tcttcattga	ttgggaatga	tgaagataaa	ttcatgatta	58260
tgaatactgc	tggatatggc	tatatattcag	agttttaaaaa	tatggtttct	aataagaaat	58320
caggaaaagc	atztatgaaa	atccctcatg	aagcagacct	tcttaaagcc	attaaagtaa	58380
gagacgatca	tttgtatata	gcagcagttt	caaataattgg	cagactttta	atttttaaga	58440
ttgatgaatt	gccaaactctt	ggaaagggca	aaggggaataa	aataataaat	ataccaaccg	58500
ctaaatttat	agcaaaaagaa	gagttaatga	cccatgcgca	acttgtttct	gaggctagct	58560
ctttaaggat	tgaaagcgga	aagagattcc	tcactttaaa	actcaaagat	ctagaaaact	58620
atatttctac	aagagcaaaa	aggggaaata	tgcttccaca	aggatatagg	aaggtagata	58680
aatgattga	agaggttgag	ttagaagtta	aagaagactg	attatagatt	tctcaaaaat	58740
ttttaatccc	tcatcaacta	gctcattttc	tattattagg	cttgagaaa	acctaacagt	58800
agatgcattc	gccttaagaa	tcattaaccc	attatgatgg	gattttttta	ttaaatcatc	58860
tatttgaatt	ttgctatcct	tgcttacttc	cactccgacc	cataaaccag	cagaagttat	58920
tttttcaaaa	catttgtgct	tctcatttat	tttattcaac	aaattgataa	atctaacttc	58980
tttctttaaa	accttattta	aaaatgattt	tttagaaatg	gtatctataa	cttcattacc	59040
tatggcacat	gcaattgggc	ctcctccaaa	agtcgtgcca	tgagaccctg	ccgacatatg	59100
ttttgaaact	ttatctgaag	taagaatgcc	tcctagtggg	aagccattag	atataccttt	59160
agcaaaaaca	agaatatcag	gagtaatggt	aaattgctca	taagcaaaga	gtgttccagt	59220
cctgccaatc	ccggattgaa	cttcatcaat	tatgacaagt	gccttatggt	tttttgcaag	59280
tttctttatt	tttgcataaa	attttttata	tgcttttggt	attccagatt	gccattgaac	59340
cagttctaata	ataaccgctg	cagtcttata	tgaaaaaacc	ttttctagat	tagtaatgtc	59400
attatatgga	tgattttttta	tcctcttagg	cagaggcgca	aaaccgtctg	taagatgctt	59460

agcttttagct	aaagctatac	caagcatagt	tcttccatga	aatgacgttg	aaaatgaaat	59520
gacctcattt	ttatTTTTt	taactgttga	tgagcaaaat	tttcttgcta	tcttaactgc	59580
agcctcaatt	gattctgccc	ctgaattgca	aaaaaatact	ttatctgcaa	aagaattctt	59640
gcataacttt	cttgccaaag	taactgaggg	ctcatttatg	tagagatttg	ataaatgcca	59700
aagctcctca	gattgctttt	taagaatttt	tattagatct	ttatttgaat	gaccaagggt	59760
agtgacggca	atgcctgctg	tgaaatcaat	atatttctta	ttatttaggt	cccatacatg	59820
cgatccactc	gctTTTTt	caacaaaatc	tgcaggagcg	taaaaaggca	ccatataatt	59880
agtatatctt	ttaattatct	ctttcataaa	tatattgtat	cgtggtttca	gttcaaaaat	59940
taaatgacca	aaaatttcta	gtatctttta	aaccaaacag	ctcactcatt	ggctttaata	60000
gaataatatt	tatatctagt	atatctattg	tatgcggtgg	catagcattg	atatttttct	60060
tttttggagc	aacgctcatt	ttgccttttg	ctggtttaga	gcttggtatt	ctattcactg	60120
cattctattt	aagttttaaa	tggagtgata	aaaaagaaaa	aatattttatt	tctcaagatc	60180
ttgtaactat	agaaaaaggc	tctaattatg	ctgaatataa	atgggaagag	ttcagggtcat	60240
ttacctcttt	ccagggtttca	aaagacagaa	gagatcttct	taagctaagc	tttaggtcta	60300
aaggcgagga	tgttgagggt	ggcagctttc	taaatgaaga	tgataaaaat	gtattaatag	60360
aagagctaac	tcagatcata	gatacattaa	atcacgattc	cttctcaaag	ccagagcttt	60420
aatttctttt	taacttcttt	tagctttaat	tcttttattt	cagggatgcc	atttctaaaa	60480
tttggaatg	attctccctg	tattaatggc	ataagatatg	caattgcttt	tggggtaacg	60540
tccattccat	ttttagctat	aaaagatttt	ggtaattttt	tttctaaatt	agctattttt	60600
gccagtggag	ctggttcaat	tttccaccta	tatttttttag	ctttacctct	aactattatg	60660
ggcataacac	cattcatgcc	ttctactgca	tactgaactg	ccttggcccc	aacagccata	60720
gcttgctcta	ggtctgtttt	ggaggcaata	tgtcgtgcac	ttctttgtag	atagtcagaa	60780
acagcccaat	gatttttttag	ttttaattta	tcagtaatta	agttggcaat	gtaaggcgca	60840
actccaccaa	gttgagcatg	accaaagtgc	tctattgttg	ctgattcaga	aagaaatctg	60900

ttcttattat	ttttcaaccc	ttcggatacc	acaacaacac	agtagccatt	ttttttaaca	60960
acacttttaa	cttctgctag	aaatTTTTTT	tgattgaagg	ttatttcagg	taataaaatt	61020
atatgaggtg	catctccttt	ttcttttctt	gcaagagaag	atgcagccgc	catccagcca	61080
gcatgcctgc	ccataacttc	taatataaaa	acttttggtg	atgttgctgc	cattgaccga	61140
acatctaattg	aggcttctag	agttgatgtt	gctatatatt	tagccgccga	cccaaattcca	61200
ggacaacagt	ctgttaatac	caaatcgtta	tcaacagttt	ttgggatggc	tatacagtta	61260
attggataat	ttaatTTTTT	acttatctgc	gaaactttta	atgcagtatc	agctgaatca	61320
ttaccgccat	tgtaaaaaaa	atatacctata	ttgtgcgctt	taaaaacatc	aataagcctt	61380
ttgtactctc	tttcgcttga	ctcaagatct	tttaacttaa	acctacatga	cccaaaagcc	61440
ccaccaggcc	tatatTTTcaa	tgactctaatt	gcagatatgtg	attcttttga	tgtatctatt	61500
agttcctcat	tcagcgcgcc	caatattcca	ttcttgcttg	cataaatttt	accgatgtct	61560
ttatgcttct	tggectctaa	aatcagagcg	ccagctgtag	catttataac	tgcggtaacg	61620
ccgccggatt	gagcataaaa	cgcattTTTT	ttcattTTTT	ctccatcaaa	tgtactgaaa	61680
taaaaatatg	taaagtatta	taactaatat	gcgaatacat	atcttaggga	tctgtggrac	61740
ctttatgggc	ggccttgcta	agataacttaa	agagtcaggg	catgaaatat	ctggatcaga	61800
tattcaattt	tatcctccca	tgtcagacta	tcttgatagc	tttgatattg	aatgatcaa	61860
gggctatgac	ataaaaagca	tgccagatgc	tgacttgtat	gtgattggta	atgcactttc	61920
tagaggaaat	gaatccggtg	agcatatTTT	gtccaattct	cttccgttta	agtcaggccc	61980
tgaaatgctt	ggagaaattt	taaaaaataa	aaaggTTTTa	gctatatctg	gaacacatgg	62040
taaaacaacg	acttcatata	tgttgactca	tattatgtta	gatcagggta	gggatgtcgg	62100
ttttctagtt	ggtggaatat	caaataatat	atctggctct	gcatgtcttg	ggtcagatgg	62160
aacttttgtg	attgaggcag	atgaatatga	ttctgcattt	tttgataaaa	gatcaaaatt	62220
tattcactat	tcaccaagca	ctatagtcac	caataatata	gaatttgatc	atgctgatata	62280
ttttaataat	ttagatgaca	ttaaaagaca	attccatcat	ctaattaaaa	taattccatc	62340

aaatggaaat	gttggtttatt	ttgctgatga	taaaaatata	agagatctta	tcgatatggg	62400
aatttggtca	aatcagatag	caatcaataa	taatgctcat	tcaattgagg	cagttttattc	62460
tgataagact	ctaaaatatg	aagaaagtat	ttattcatta	aatgagttac	ctttaatagg	62520
agaacataat	tttaaaaact	acatttcggc	tattttggcg	gcaaagacag	atggaattca	62580
aattcaggat	tctatcaa	cattagctag	ctttgatggc	gtaaaaagaa	gattagaata	62640
taaaggaagt	tttgatggca	tagaaattta	tgatgatttt	gctcatcatc	caaccgcaat	62700
agaattttct	tctaattgcc	tcgtaacaca	aatccatca	aaaaaaatac	ttgggtctcat	62760
tgagctaggc	tcaaacacta	tgtctggggg	ttctcatggc	ttgtctttgg	ttgaatctgc	62820
aaaatcttta	actcatgtta	tctggctaga	tcgcaataat	gttttgtcag	agaatgctag	62880
cattgaatct	actaacacta	ctgaagattt	tattttctgca	gcgatatctg	ctttctcaga	62940
ttatgatatt	gttattttga	tgaccaataa	agacagccaa	aaaatattaa	aaccattgt	63000
agatcacttt	gaaaaataat	aatttaccag	tttttccttt	aggaatagtc	gccctcccag	63060
gtagcatcca	atctcttcaa	atttttgagc	ctaggtatat	acagatggtt	aaaacatgtc	63120
tatccaagaa	ccatggattt	gtaattgttt	ttaatgccaa	taatgagtct	caaggcgatt	63180
tcactttttc	taagaaggga	agttttgttg	aaattataga	ctttaataat	ttgccaaatg	63240
gccttcttgg	gataactgta	aatctataa	ataaggatgat	aatcagtaat	atatgtcaat	63300
tagaagatgg	actgcatatt	gctgatatta	aggcacagat	agatccagag	gtagatgatc	63360
aagctgtttt	ggcagaatat	cctgagatat	ctagcattct	ttctcagctt	gtaaagcatc	63420
ctaagattag	tgacctgccc	atccagggtg	actttggctc	tgctgattca	gttgcatacc	63480
acttagcagg	ccttatacct	ctaagctcaa	atgagaaaca	aaaactatta	gaagcattcg	63540
atgcagcaca	gcggatgaga	attctttcag	actatattga	aagaatatct	actacataaa	63600
ttattttatt	taatattatt	ggcggcttat	tagattttgc	tctaagccaa	ttgatggact	63660
taaaaaccgt	agggatttct	acaaaccttt	tttcaatata	gtatttgcct	ggatagttac	63720
taagcattag	ccaataata	atagtgaaca	aacctggcca	ggtaacacca	gcatcatgat	63780

tcccccaata	ataagggaat	aacctaaaat	atttttaatc	agtaaaacta	aataccacaa	63840
caacgggtta	ttggtcttaa	atttcgagac	atctttcttt	ataaaatagt	cacttggaat	63900
taatccagca	agccacctca	tactaacgag	actaaataaa	aaaataaata	aagatattga	63960
gctcagccaa	agtataagaa	ctttataaga	ttcaaaaaaa	attatgatgg	catttagggg	64020
attaaattcc	ataattaccc	acctttttat	ttaataatga	tttagtatag	caatgtctaa	64080
aaataaat	aaacattaaa	taaatataca	ataattttgc	caaaaaaagc	cttaaattgat	64140
atttgcactg	aaaccctcag	tgataataaa	gccgaaaatg	ttctatcact	tgatatcaaa	64200
ggatatttctt	cttttgctga	caacatcatt	attgcaactg	caaattcgaa	taggcatgca	64260
aagtctttat	ccgaaaagtt	agttgaggag	ataaaagcta	ataaaattag	tatcatgggt	64320
gtagagggca	agacagaatc	aggttggata	ttagtagact	gtggtgaggt	tgtagtaa	64380
attatgaaga	atgacataag	agagttttat	gatttagaag	gtctttgggg	tgaaaacacc	64440
ctcatcgatt	cttcgaaata	aatgctatta	aatatcataa	gtggttgaaa	ctcaccttca	64500
tcttgaggagt	taacaggaat	agaatattac	acaaagcaaa	tccctaaaga	agtaagtcta	64560
aattttatta	acgtaaaagg	gcagcaacac	ccaaaaagat	ctacagaaga	ggctcttaaaa	64620
cttgaatcta	agttaatctc	gtccaaaata	gattctaattg	gatataattgt	ttgctgggac	64680
tcaagtggcg	agtcgttgaa	taattttgaa	tttagtaa	tttttgaaaa	atccatgctg	64740
gaaaatatga	agctttactt	tgtaattggg	ggttcatttg	gcatacctcg	agatatctt	64800
gataaatcta	ataagataat	ctcgatctca	agtctcacc	tacctcacag	gcttttcaag	64860
attgttctta	tagagcaaat	ctatagggtca	ttttcaattc	tcaaaaatct	tccctaccac	64920
aatgattga	tctgaatgag	aggtttggtg	aaaaaagaag	ctttttcaat	agattggtgc	64980
ttatatattt	tttctttggg	gccctgtttt	tattcttgct	tttcaagacc	tattctttac	65040
aggtttctag	ctattctgac	tatgaactag	cagcactaaa	gaataaaaca	aaagaagtgc	65100
tagttcagcc	tgttcgagga	gtcatctatg	acaggaatgg	aaatattcta	gttaataatg	65160
tccctacata	tgacctaata	atccaacat	caaaaattaa	aaacttagat	gatttcatta	65220

ttaatatattc taaaataata gagctatcag attcagaaat tgaaaacatt attaaaaatt 65280
ttaaaggag tgcaacttat aatagagagc taacaattaa aaaagatctc tcgaaagaag 65340
aaatagcaaa atttgaagta agaagctacc aatttcagaa tgcatttata gatgttaggt 65400
atagtagaga aaataaatat ccttacttat tttcacacgc actcggttat gtaggtggtg 65460
taagtaatga taaggtttta tcaattttaa aaaatcaaga tttgaagcaa tctgaaacaa 65520
cttttaagta ttcagggtggt tttattgctg gaaaaacagg attggaaaat atttatgatt 65580
cagctttaag agggtccttt ggaaaaaagt tatTTgaagt tgatgcaagg ggaaggcttt 65640
taaaagagct aagttttgaa aaacctatta atggaaaaag cctttttact catttagatt 65700
taaattcaca aaaaaagcc tttgaacaaa tgaataatag aaggggtgct gttggtgcct 65760
tagagcttaa atctggttct atagttacct acctaagcac tccaagcttt aatgtaaattg 65820
gtctttctaa tggcatgtcg tcggtagagt tcgaaaaact gattaatgat gtagacaagc 65880
cattttttga tagggccggt caaggtcggt actctccagc ttctacaata aagccagcaa 65940
tagcattggt tggattataa gaaaaaatag tagactggaa ctttacactt aaagatcctg 66000
gattttttgt attaccagag gatcagagga tttatagggg atggaagaaa ggagggtcatg 66060
gaacaattga tatgaagggt gcaatcatag aaagttctaa tacttttttc ttttctcttg 66120
cttataaggc tgatattaat aatttaatca gccatctttc tgagtttggt tttggcagga 66180
atgtttgtaa agattgtttt aatccagatt ctggattatt gcctacgcct gaatggaaaa 66240
tgaataatct taattttggc tggtttaaag gagataccgt taatttaggt gttggtcagg 66300
gctatatgag tgccactcca gttcaattag catattactc tgcatttctg gcaaaaaaag 66360
gaaatcttca agagctatca tttgtcgaga gtgacagcct aagcaatact gcttttataa 66420
ataatttaag catagataac tcagactggg atcaaattca ctcaagcatg attggagtta 66480
ttgaagatcc gagaggcact gctaaaagat taaaaccttt aaaatcatat gttggtgctg 66540
caaatcttg aacggttgag cttgtcagca cacaacaaa ggaagattac aaactagtaa 66600
ggcaaaatat tggtaacaga gaccatgcaa tcattgtggc ttttgggtcca atgcctgac 66660

cagagtatgc agtaagtgtt gttattgaaa atggtgaaag cggcgggttct gttgcaggtc	66720
ctgttgctat tgctgtttta aatgagctta tcaataaatg aagaaaaaat tagacttta	66780
aaacttttagc atttattttg atcaatatatt atttattgcg ataaccctgt tgcggtcat	66840
gggtttatatt tttttataca gcgcatacaca agaagatatc agcactgttg ctaagcaagc	66900
tgtattcggtt gggttttggtt tgctgttaaat gttttagtatt agccaacctg accctgattt	66960
ttataatata ttttctgggt tatttttttg ggggggastt gtattgattt ttctaactat	67020
gatttttggt aaagaaataa atggagccaa aagatggctt gatttaggat tttttaccct	67080
gcaatcctct gagattatta aaatatcatt gccaatTTTT ttatcatcat atttatataa	67140
taagccactt ccaataagca ctaggcatac ttttattaca ttgatattaa taggttttat	67200
atttgcactt gtagctagac aacctgactt gggaaccagt ttagtagtat ttatgtcagg	67260
gggttatgta ttatttttag caggattaag ccgccgttta ttggatctgc aatagcctta	67320
tttttattat ccttacctbt tttatggaat aactttctag agccatttca acaacagaga	67380
gttttaactt atttgaacc agacgcagac ctttatggta ctgcatggaa tataactcaa	67440
tcaaaaatag caattggatc aggcggaatt aatggaaaag gttatcagga gggctcccaa	67500
gcccatcttg atdttttacc agagacagaa acagatttta tttttgctgt tattgctgaa	67560
gagtttggtt ttattggagt ctgtattttg ttatcagtat ttdtctttat atkactcaga	67620
tgtttatatt tagcatttaa tgcaagagat agattttgca ggttaactat aggaggccta	67680
agtttagttt ttgcctctac attatttatt aatttagcaa tggttggttg ttagttcct	67740
gtagttggta tgctcttcc atttatcagc aaaggcggct catctttgct atcctkttat	67800
atagcttttg ggattataat ttctatggca acacataaaa aattaatgca aagatgaaaa	67860
aaattatatt tataacttta atattcacta tttcaattac tgcggattat tcgaatcatg	67920
aagatagcca aatggtaata aatgaactcg tcacaaaaca cggttttgag gaatcttatg	67980
ttactgcaat cttaaaaaat gcaaaaaagc gtgatgagat gcttaaactct gttgctaata	68040
ctgctgaaaa aacaaaaaca tgggatgaat atagggtat ctttataaaa acaaaaagag	68100

tttcagaagg	caaaaaat	ataaaaaaa	atattaatgc	tttagagagg	gctgagaaag	68160
aatttggagt	tcctaaagaa	ataattactg	ctattttagg	cgtagaaact	aattatggca	68220
gcaataaagg	tggatacaga	gttttagata	gtcttactac	tttaggtttt	gatgacccgc	68280
gaagatctaa	tttttttaga	agagaactta	tagagttctt	ccttttaaca	agagaaaata	68340
atttagatat	caaaacaaca	aaaggctcat	atgcggggagc	catgggatat	gcacaattta	68400
tttcgtcaag	ctaccgagcc	tatgctgtag	attttgatga	agatgggttac	gttgatttat	68460
ttaattctgt	tgatgatgca	attgggagta	ttgcaaatta	tctttatggt	catgggatgga	68520
agagagaagg	aaagatcgta	acaaaaactt	atccaaacaa	tgttagaaaa	ttttataaac	68580
ctcatgagtc	tctaacaagg	ttcatacctt	taatctttta	tgaagatgga	aaagatcttt	68640
tttttattgg	tgatgataac	tttagggcta	ttgctaagta	taatattagc	gatgtctatg	68700
caatggcggt	ttattactta	tcagaagagt	taaaaaatg	aaaaaattat	tattcacatt	68760
actatcaact	tcaatattta	ttcaagcaca	gagttttggt	ccagattctc	ctgagttaga	68820
tctgaagagc	tatatcctaa	ttgagccaaa	caccaatact	gttattgcag	aatttaattc	68880
agatttggaa	atagaaccag	ctagcatgac	taaaattatg	actagctatg	ttgtggctga	68940
tcagattgca	aatgggttta	tatctcttga	tgacgaagta	ctaattagtg	aaaaagcatg	69000
gagaatgcaa	gggtctaaga	cctttataga	agctggaaaa	aaagttacgg	tatctgatct	69060
tttaaaaggt	attatgattc	agtcaggcaa	tgatgcctct	gttgcaattg	cagaatatgc	69120
aggcgggcact	gaaagaggct	tcgttgactt	aatgaactct	tatgctgcct	ctttggagat	69180
gaataatacg	atctttcaga	attcaacagg	gcttccagat	gaaaatcatt	tttcatcagc	69240
aaaagattta	gccaatctga	cagctaacta	cattaataaa	tttccagaag	aatatgcttt	69300
atataagcag	aagcaattta	cttttaacaa	tattaagcaa	ttaaatagaa	ataagctttt	69360
atggagagac	gactcagcag	atggtgttaa	aactgggtcat	acagaagctg	ctggatactg	69420
tttagttggc	tcagcaaaaa	ggggcggtat	gaggcttatt	acagttgttg	caggaagtaa	69480
gtcagataat	gatcgttttt	tatcttccca	aagattactt	gagtatggat	ttagattttt	69540

tacaacacag	aagatgttaa	gtgcagaaaa	agaatatcag	aatattacag	tttggggtgg	69600
acaagaaaag	atacttggcg	tgggagttct	agaggatata	tctattactc	ttcctagaac	69660
aagtttcaaa	aatgttgaaa	ctgtttataa	agttaataac	aatatccagg	ctcctatcgt	69720
agttgggtcaa	aaagtcggga	ctctagaaat	aattagtaat	gatgagattg	ttctagttac	69780
ggatttggtt	gctttaaaaa	atattgaagc	taaaggtttt	tttggaagaa	tatgggtcaaa	69840
gttcgttctt	tggatattta	gcttatttgg	tctaacagat	gaaaacacca	cttaaaggcg	69900
tttttaatgg	ggcttttgat	actgtagata	acataaaaat	atcacctttt	tctcgagctt	69960
atactttttc	tgatagcgta	tatgaagtag	ttcctttctt	taattcaact	gcaattgctt	70020
ttgatgatca	tataaaaaga	ttagaatttt	ctgctagtca	actagcgatg	gccgtagatt	70080
taaaagaagt	tatatattgaa	attaattctt	taataacatc	atctgaattt	agtaatggct	70140
atgtctatta	tcaagttact	cgaggcggtg	atccattaag	gtctcatatt	catgagccaa	70200
atttaaaaat	agaaaccttt	ggctatgcaa	aggcgacttt	attccaatgg	aaaccattaa	70260
gagtatcagt	atgtgatgat	attaggtggg	ggaggtgtga	tattaaatca	acatcattgc	70320
ttggtaatgt	tatgaacatg	aatgcagcaa	aactagataa	ttgtgatgaa	gttattatgc	70380
acaagataa	tttgttgacc	gagggtggcg	catctaattc	tttttttgta	aaaaatgatt	70440
caatctgcac	cccagctcta	aatggaaata	tacttcccgg	cataaccaga	gcattactaa	70500
ttaatgaatt	aaaaaattat	agtatagagg	tcacgaaga	taacttcagg	ctagaagatt	70560
tatcaaattgc	ttcttgcgcg	tggctcacia	gcgcaacgaa	agggctggcc	cccatttctg	70620
aaattagtaa	cctagaatct	cacttggatt	tataccatcc	tcttttcaaa	aagagtgaag	70680
aaatttttaa	taagaagttc	cttagttaga	agtcagttta	ataactatat	tatcaagctc	70740
gtaccaaaaa	tctttatcac	taagaccttt	atttgcaaga	tcaagctcaa	atacattttt	70800
ttggagctgt	attaagtttt	gtagtttatg	atttttaata	aatgcatat	attcagaagt	70860
tttatttttc	caaaccacag	atttaattag	actatctaca	ggatttttac	tttgtctagc	70920
attggctgca	ctgttaataa	ttttaccaac	tatccaaact	agaagaggtg	cataatgatc	70980

ttccgctgat	tttattgaat	ggataatfff	taaagcttgc	ttggtattat	ttgaaataat	71040
tttatcttca	agctcaaagt	gcataaatfc	tgcagactcc	acagaagggt	ttctttcttg	71100
gccatcacca	tctttataag	tcagtcttag	aagctttact	tcattctggt	gggcaactaa	71160
gtttccagag	ttcatttcag	atatatfff	cactaaagag	ggcctatctt	tttcagaaat	71220
aaaagatagt	tgatgcttaa	gccaaatctt	ttcttcaaaa	gatttaaggt	ttccacaatc	71280
aatgatgagg	gctatctcat	ccatctgttt	taccactta	gttgtgaaat	taagcttttc	71340
attgcttgaa	ttaataatta	ttgcaatatt	atccatatgg	ttaatatfff	caatttgaga	71400
tattttaatt	atttggtcag	gtatfffcc	ttgatcatga	ttaatctcaa	ctattatfff	71460
tgagccaaat	aaagatccag	aagcattttc	aataatagtc	tgatctatff	tatcaaacc	71520
atctttatta	ataattgttt	tttctttgaa	ccctttatff	gataaatggt	tgagcaaaa	71580
atcttttagag	ttatfff	agacaacctc	agatccaaaa	ataaagaaga	tattttgtga	71640
ctgatctaaa	tatttttgag	ttgtcattgc	ctcacatttc	aaaagcatgc	acctcaaaaa	71700
ttatttgatc	aattaattcc	ttctgcattt	cagactcaag	agatttaatc	atttgatctt	71760
gtgcaaattgg	attaagttca	ttcgatttat	atctcttgta	gctaacaagc	tttttactta	71820
tttgttttatt	tccaatcatt	tttatatgta	tttctaaatt	aatttctcct	tcaagagctc	71880
tttagagaaga	gcctccatag	atatcatatt	tatttaaaat	atagtcttgt	attagaatct	71940
gattttgacg	attctcagct	tctttttctg	tattgaagag	gattgctaac	ttttgctcaa	72000
aagaattagg	gacactacta	tcaaaccgaa	aattaaatag	atctttttta	tctagatcta	72060
tatttagttg	attaaattgg	cagccagtta	aagatatcaa	caagcaagtg	tataaaaaga	72120
aaggctcttag	tttcataacc	tcatagtatc	tatctgaaga	atatatatca	aagattaatt	72180
ttaaataaca	aaatttatta	ttttttcttt	aatataaata	gtttttttta	tagccgagtt	72240
gtttattggt	gatgccacat	tgtcgatagc	cagtgcataa	gcttctatat	ctttttgctc	72300
aaggtttttg	tctatcatta	ccttacctct	cacctttcca	tttacttgaa	ctattaaatt	72360
aaattcagaa	acttctaata	actcctctct	aagaacaggc	caagaagatt	caatttcttc	72420

ttgtgcaaag	tcaaagtaaa	aattattcca	taaatgttgt	gagatatgag	gcgcaatagg	72480
attgagagtt	ttaagaataa	ttattattgc	ttcattaaga	caatattgat	ttgagatcga	72540
cgcattacct	tcttttaaaag	aatcaggtat	aaagttaact	agctccatta	tcgaagcaat	72600
ggctgtatta	aatgaatatc	ttgtttcaaa	gtcatttgta	acctttttta	gagtattatg	72660
cgatttttctt	cggagctcta	cttcttggtt	cgaaggatcc	ttgggctcct	ctaagtcaat	72720
aaatttttcta	ttgctgacta	agttccatat	ttttttcata	aatcttgagg	caccttctac	72780
agatgactct	gaccattcta	gactttgctc	agggggagct	gtaaacatca	tgtaaagcct	72840
tacagtgtca	gcgccatact	tttctatata	agattgagga	tctacagtat	ttccttttga	72900
tttggacatt	ttagccccat	ctttaagaac	catgccttgt	gtaagaagct	ttttaaaagg	72960
ctcattgcct	tcaactaaat	ccatatctct	aagcgcctta	tgaaaaaatc	ttgaatatag	73020
taagtgcagt	atcgcatgct	caatacctcc	tatatataaa	tcaacaggca	gccaatattt	73080
tgaattttta	tcaaacattt	catctgcatt	gtcagatgaa	gtaaatcttg	cgtaatacca	73140
tgatgagtcc	ataaatgtat	caaaagtatc	tgtctctctt	tttagtcgat	cagaaatatt	73200
ataaaaaatct	tcattttgac	ttaagggaat	aggtgccgag	ttctttttta	gctctggaag	73260
cactataggc	atatacctttt	catcaataac	tcttggttca	ccattttcat	aaacaaccgg	73320
aatagggcaa	ccccaatatc	tttgtctgct	cacaccccag	tctcttaatc	tgaactggat	73380
taactgctca	ccaagggttg	cgtctgctaa	atctttaata	atttcaagag	atgcctcatc	73440
agagtccatg	ccatcatatt	tatcagagtt	tattaatttg	ccttttttga	caactggaag	73500
ttcattatta	ttatcaaaac	taataacttg	aggaatctct	agattatatt	tagatgcaaa	73560
ttcgaaatct	ctttgatcat	gagcaggaac	acccatcact	actccagttc	cataatcaag	73620
aagtacaaaa	tttccaatcc	atacgggtat	cttcttcttt	gtaattggat	gaattacatg	73680
cattccacta	ttaattccaa	gtttttcagc	tttagccata	tcagcttcag	cagctttcac	73740
ttctttgcat	ttgttttagaa	agtcttttat	cgattcatta	ttttttgaca	accctattga	73800
tatagagtga	tttggtgata	ttgctaataaa	agaaacacca	aaaattgtat	caggctcgtgt	73860

tgaaaatacc	tttaaagaat	catcggaatc	ctcaatcatg	tatttaattt	ctgctccctg	73920
agatttccca	atccaattcc	tttgcatgt	tttaacattt	tcaggccaat	ccacctcatc	73980
caaagatggt	aaaagttcct	cagcgtagtc	tgtaattttt	ataaaccatt	gatcaatttc	74040
ttttatttca	acttgcgccc	cagacctcca	tccttttcca	tctataactt	gctcatttgc	74100
taaaactgtc	tcatcaactg	gatcccagtt	cactaatgat	tttttcctat	aaaccaaacc	74160
tttatcataa	aactttttta	atattaactg	ctcccactta	tagtattccg	gttcacatgt	74220
tctaagctct	tttgaccaat	catacccaag	gcccaaagat	ataagttgct	gtttcatatg	74280
ttcaatattt	tgattcgtcc	aatcttttgg	gctaacttta	ttagcaatcg	ctgcgttttc	74340
agcaggaagc	ccaaatgcat	cccagcccat	aggctgaaaa	acattaaagt	tattcattct	74400
tttatatcta	gatatgacat	caccgatcgt	atagtttctt	acatggccca	tatgtaattt	74460
acctgatgga	tatggaaaca	ttgataggca	ataaaatttt	tctctaccat	caggattagc	74520
tttaaatttg	tcttctttta	accattcatc	ttgaatggtt	ttctctattt	gacttggggt	74580
atattcagga	ttcatttttt	taggaactca	ttttctaagt	aattaatagt	atgtttattt	74640
tctttaaagg	tatcatcatg	cagaatcctt	tgatgaagag	ggtgatttgt	tttaatgccc	74700
tctataaaaa	attcatctaa	agcactcagc	attctcttaa	tagccgagct	tctagaattt	74760
gcctgagtaa	tgatttttgc	tagaagtgaa	tcatagttag	ggggaactct	atatactcca	74820
taaatatgtg	agtcatatct	tatgccaaaa	ccgccagggtg	tatgcatttt	tgtaattgtt	74880
cctgggggatg	gttgaaaatt	atcaggatct	tctgcattaa	ttctgcattc	tagtgagtgg	74940
ccatgaaaat	taatatcatc	ttgatttagc	tcaattggca	tttcaagagc	aattcttagt	75000
tgtgctttta	ctaaatcaaa	gccagttatc	atttctgtta	ccggatgttc	aacctgaatt	75060
cttgatttca	tttctataaa	atagaattga	ttatcttcgt	ataaaaattc	aatagtgcca	75120
accccttcgt	aatttatattc	ttcacataaa	ttgacacaag	ctattagtgt	tttattaaga	75180
gcttcttgat	ttatatthaag	tgctggagct	tcctcaatta	ttttttgatg	tctcctctgc	75240
atactgcaat	cccttgtacc	taaatgtatc	gcctttcctt	taccatcacc	aactatttga	75300

acttcaatat	gtctaggggt	cccaataaat	ttttctagat	aaatggtttc	attgccgaat	75360
ccatttttag	cttcttgc	tggtatttct	gcatgtccaa	taagatcttc	ctctctttca	75420
acaactctca	ttctcttcc	accgccacca	gccgttgcct	taatcataat	tggatagccg	75480
atatctctag	caattttttt	aaactcatca	ttatcagatg	gaatctcatc	tttataacct	75540
ggaacgattt	gaattccaga	tttttcagcc	agtgtttttg	ctgttatttt	gtcgcccatt	75600
ttttggattg	tttctgaggt	tggcccaata	aatttaaata	cacttttttc	acacatttct	75660
gcaaaattat	gatcttctgc	aaggaacca	tatccaggat	atatagcatt	agcacttgta	75720
agttctgcag	ctgagagaat	ggcaggtata	tttaggtagc	tttgagttgg	agatgcaggc	75780
ccaatacaaa	cagtttcatc	agagaatctt	aaatgtttta	gatctttatc	agcttctgag	75840
taaacagata	ctgtcttgat	accaagctct	ttacaagctc	ttatagctcg	aagagcaatc	75900
tcaccacgat	tggcaataag	aactttataa	ctcattttta	ttgacagtta	taattttttg	75960
tccaaattca	accggacttc	catcttccac	gtcaataact	gaaatcttgc	catcaaattc	76020
tgatttaatt	tcattcatca	ttttcatggc	ttcaactata	cagagcacat	cacctacttt	76080
tatatgtct	cctactttta	caaattggatc	tttttcggga	cttgggcttc	tataaaaagt	76140
tcctactatt	ggagaggtaa	caacatcccc	tatgacagtt	tgcttagcct	catccaattg	76200
atgagttggt	ggtgagatag	ctggctcatt	tttgacaatg	attggttttg	aaacaaattc	76260
ttgattagag	ccgttatctc	ttgaaattct	tactgactca	tctccttgac	ttacttctat	76320
ttcttttaaa	tcagactctt	gaagcatttc	tataagtttt	tttatttttc	taatatccat	76380
ttattttacc	ctgtatttta	atatgccctt	aagcatcgcc	tcttcataac	ccttagcacc	76440
aataccagtt	atcacttctt	cagctatata	agatagatat	gagacatgtc	taaattcctc	76500
tcttgtatat	atgtctgaaa	ggtgaacctc	atagaatggt	atgttgacac	caagaaatgc	76560
atcccttata	gctatgctcg	tatgagtata	tgcagcagga	tttattatta	caaagtctat	76620
tttttcttca	atagtgctat	gtattgattc	tattatttca	tgtctccgat	tactttgaaa	76680
agatattaaa	ttgcattgat	tttgatttgc	taagactaat	aaatcttggt	gaatatcttc	76740

aagtgatttg ctgccatata cttccttttc tctagttccg agaagattga ggttggggcc 76800
atttataagc aatatattca tataatagat tttaacagaa atttacagat tttttaggat 76860
tttaaagata tttttataag agttgatttt tttgcatggg atagcaaaaa cccgcttcag 76920
agcatccctg atagtatatt aaaacctccg ataagttcaa tgaagaattt atttcaatgc 76980
ttataataaa ttcattctta agaattctcag tctcgccaaa aaattcatcc ctatatagat 77040
ctttacttga ctcaagcgtc ttaaattggga gcactttatc tttaaatttg agctctatac 77100
tatccaagta catataataa ccatctctta tttgccaagt tatatatgct gcttggtttt 77160
caacattggg tgtaaagca aatacttcat tggcttctgg gaccctattt gagttttcaa 77220
aaagatttgc tgaattcaat tcacctgcaa tcacattgga tagtagtata attattaaac 77280
ttttgatcat gaagacttaa gaataaaccc ctaagtactt aaaataaaga ttttattaaa 77340
taacacggta gttaaaaatt tatgttccag actttcgcac taaaaatttt cttttatatg 77400
cctatttggc ttttaaaaat aattttttat agaaaaagaa cagtaataag agggcatcaa 77460
tttgatgctc aatctgctgc attattgtct ttgcttccaa agaaagattt atctgaatta 77520
ttagatggg aaattgcaaa agctagaatt actcttgaag aagcaagaat tcaaaataaa 77580
gtttctttta caccgtctat acaagtcaga aaagtagatc atattttacc aaagcatgat 77640
ttaattctca gagaatacaa gccgcatcaa gaggatttaa aaaaagttat tctctatttt 77700
catggaggag gttatgtcct taactcagtc aatacacatg atgatatggg ttcatatatg 77760
tcagataaaa taggagtaaa gttttattct ctagactata gactatcgcc tgagagcaaa 77820
tatcctgatt ctcttgatga tgctctagat gctttttctt ggctcatcgg ccagggatat 77880
ggaccaggtg atatttcagt ttgtggagac agcgcaggcg ctcatctagc tgcttctttg 77940
tgccattacc tagctgaaca aaataaagat ttgcctagta gccagctact aatatatcca 78000
atgtgtgac cgtcctgttc atctgagtct tataatttat tatcttcagg atatcttcta 78060
actaaaaaaaa ctatgatttg gttttgggat aagcttagaa cttccgaagt aaacaataac 78120
gattcggcct ttaattctctt aaaatttaat ttgaaaaaa ctttaccgcc aacaattatc 78180

gttacaggtg gctttgatcc tctatgcat gatggagaaa aatatgcata tttattacat	78240
aaaggtaaac ataatgtgaa acaattacat tatccaacaa tgtttcatgg gtttgcacat	78300
atgactaaat tgaaagcagc gcagatagcc gttgaagatt ttttaaaaga atataagaaa	78360
atactatgag taaaatttta gaagtaagtg acttaagcat caatttttca acaagggatg	78420
gattgtttta tgcagtgat aatataagtt ttgatataga aaaaaatcaa accttggcct	78480
tggttggtga gtcaggttct ggtaagtcgg taactgctat gtcaattctt cagctccttc	78540
aaaaaccaca agcatcatat tccaaagagt cttctattaa gtttaatggc gatgagataa	78600
taaatgccaa gtatgaaaag ttactttcct tgagaggaaa tattatatct atgatatttc	78660
aagagccgat gacctcacta aacccttctc acagagtagg taatcagata actgaatcaa	78720
tactacttca ctcaaaaagc tcaaaaaaag atgcaataga tgaagcaaaa aaattaatgg	78780
cacttggtga gattgatgat gttgaaagac ggttctatgc ataccctcat gagctttctg	78840
gagggcagcg acaaagagtt atgattgcta tggcccttgt taataaacct gagctattga	78900
ttgctgacga gccacaacc gctcttgatg taactatcca agcccagata ttagatctca	78960
tgtctaagct taaaaatgaa ctaggcatgt caatactttt tattactcat gatctaggcc	79020
tcgttcaaga attttctgac aatgtttgtg ttatgaagaa tggcaagata gttgaacaag	79080
gaaatactgt tgaggtattt aacaaccct ctcacgaata tacaaaaaaa ctttttagatg	79140
cagaacctca gcccaaatta gataaccctg taagtgatga gccaaataatt gagatcaatg	79200
acttaaatgt ttactactca ataccctcaa ctaatttttt taaaaagaat acttttcatg	79260
ctgttaaaaa tacttccttt agtatattata aaaatactac aattggcctg gttggggaat	79320
ctggatctgg aaagtcaact ttgggtaagg ctatagcaaa cttagtttct tataagggga	79380
atattaaatt tgagggaaga gatatcaact caaattccca aaaagaaaat aaagaattga	79440
aaaaaaatgt ccaaattggt tttcaagatc cttatgggtc attatcacca cgaatgacag	79500
tgggggagat agttggtgaa ggttttaggtg ttcactttta gcttacaaaa aaagaaagag	79560
acgaaaggat agataagggt ctgtcagatg tcggtatcga aatagtagct aagaataaat	79620

Top 50 "E" Sites

catgcttaaa	ctggcgattg	ataaaggatg	gcttgaaagc	gatgttatgt	tagaatcatt	82560
aataagtttt	aaaagagcgg	gagctgatgc	aattctaacg	tacgcagcta	aagaaatttc	82620
caaggagata	actaacaaat	gagcaatgtg	atagaaattc	gtgatgaaga	aagctttaat	82680
agcgacgtct	taaattcaga	aaaacctgta	ttggttgatt	tttgggctga	gtggtgtgga	82740
ccttgtaaac	agcttgcacc	aactgttgaa	acagttgcag	cagaaaaatc	agaaacatta	82800
aaggtttgca	aaatggatgt	tgattcaaat	agagagattg	ctgctaaata	tggaataagg	82860
tcaatccctt	cattaattat	atttaaaaac	ggagagcctg	caggagttga	agtaggtgct	82920
ctaaccaaac	aacaattaga	ggactttata	agtacagtag	tttaactttg	caaagacttc	82980
tttgcatatt	gcaagaaata	ggattatcat	ttcgacttct	aaggctgaaa	gccactcaaa	83040
acaaaaaccc	accttttcaa	tcataaaaac	aacaactaga	acggaaataa	ctaaatgaac	83100
cttactgaaa	ttaaaataaa	accaataaat	gaacttgtag	atatagctac	tgagcttggg	83160
cttgaggatg	ttggaaggct	gaaaaagcaa	gagataatat	ttagaatatt	taagcataag	83220
gcttctgaag	gtgttgatat	ctatggtggc	ggagttcttg	agattttaaa	tgatggtttt	83280
gggtttttgc	gatccccaga	aggctcttat	tgcgctggcg	aagatgatat	ctatgtttca	83340
ccaagccaaa	taagaaagtt	tagcctcagg	aaggagagact	cagttgctgg	gaagataagg	83400
acccctaaag	ataaagagcg	ctattttgca	ttaatccaag	ttgatactat	taatggtgaa	83460
gagccaagaa	agactaaaaa	caagattctt	tttgaaaatt	taactcctct	ttttcccaat	83520
gaaagactaa	tccttgaaca	aggaacaggg	tctaatagaag	atctttcatc	tcgaataatt	83580
gatttgattg	ctccaatagg	aaaaggtcag	cgtggactta	tcgtttctcc	acccaaggct	83640
ggtaaaacct	taatgcttca	aagcatagct	cattctatta	aaagcaataa	tccagaagta	83700
gagcttatag	ttcttttgat	tgatgaaaga	cctgaagagg	taacagagat	gtcaagaact	83760
gtaaaaggag	aggttgtagc	tagtactttt	gacgagccac	ctactcgaca	cgttcaagta	83820
gcaaatatgg	ttattgaaaa	agcaaaaaga	cttgtagaac	ataagaaaga	tgtagttatc	83880
ttattagatt	ctattactcg	tcttggaaga	gcatataact	cagttcagcc	tgcatcagga	83940

aaaatcaaga	agatatcttct	aatctatgca	gttcatttaa	gaccgaaggt	gatagattta	86880
atagtgggtga	ctgggattta	agcggcactc	cgttttttaa	aagtgccag	gctaataattt	86940
tttgtactgt	tgatcaatta	atttcatacc	acacacatac	tattgtgatt	ggtcacgtca	87000
caaactctct	tagcgatgag	aaaattaata	cactgacata	tgttgatggg	agctatgaat	87060
aaattttcaa	aaaatgtatt	ttttatttta	ataatactca	actctttttt	tcttgccctct	87120
aacatttttg	ctagccagga	agagtgtgaa	gaaaagccaa	gtgtttttat	tatctctcct	87180
caagacgggt	ttatctcaga	atctaataat	gtaaaagtct	tatttggatc	aaaaaatatt	87240
gaaataaatc	cagctggcaa	aggtgagatt	gcaaaaaata	aatgttttgc	aagcgggcat	87300
caccatcttt	tagttaatat	cgaagcattg	ccagagagct	ttattccttt	tgacaagggt	87360
tatttacatt	ttggaggagg	tcaaactgaa	acaattcttg	atcttgatcc	tggaacctat	87420
tctctccaac	taattcttgg	atcttatgtg	cataattcaa	aatgcagggt	aaataacttt	87480
aaagggtcaag	gacccttttt	atcagaaaaa	ataacaatta	cagttaatta	gagattagac	87540
cagttataac	tttatctaga	tgttcatcta	agttattagc	aattataattt	gcactagaaa	87600
taggaccaga	attatcgtca	taatcatcac	caactacatt	cacctttctt	attatcccaa	87660
caagaccatg	caacatagac	caaagagtta	tacattttaa	ggcaatcact	tcctcaggct	87720
cgtctgctag	attagcaaaa	ctttttctca	tattatcgta	tgttccattt	gcagatttaa	87780
gaagatctgg	gtaatcggca	aagttaccaa	cagctgttcc	aaacatcaaa	tcatatgtat	87840
gtgcattttt	taaaccaaac	cctatatatt	tgcttgcatt	agtaacaagt	tttttctttg	87900
taattttttt	tggattctca	aaaaaaacaa	gctcattaag	ttttttaaag	ccaacggttg	87960
caacagcggc	atatacacat	tcttttgttt	caaaatgcct	ataaggagca	gtttgagaaa	88020
caccactttc	ctttgcaagg	gatctaatac	ttagtttagt	gtaaccatct	ctatcacaaa	88080
gcctgcatgc	gcattctata	agttcttcct	ttaagtttcc	gtgatgatag	ttattcataa	88140
tttaatttta	atgttgacac	tgctaacata	tatcatgtta	ttatgtatac	accgcataca	88200
ttaagttgat	tacatattaa	agtcaaatac	agatatatga	acataacaaa	aataaataac	88260

accgtcatag ctcttttatt aggcaatgcg tttctatcaa atcttgaggc tttagaagtt 88320
cttgaggtta aaatgcttga tgaatatgct gtaactagag aatttcctgg aaagctcatt 88380
ccatcagacc agtctaagct agcatttgaa atacctggaa agataaactc tattaatgtt 88440
gatatcggag atgaggtcat cttaggggat gaacttgcct cattggatga tagagaagct 88500
ttagcacaat taaatcaatc aaaggctaaa tttgatttag ctgaacaagt actagcaaga 88560
tatataaatc tcagagcaga tgggcatatt tctattcaag atcttgataa ggctgaatct 88620
gatcaaatag tagctaagtc gcagtatgat ttttatagag ttaaatttga gcaaactaag 88680
ttactagctc cttttaatgg agttattcaa aatagatttc tagatacagg atcagtaatc 88740
aatgcagggtg tccaagtttt agaaatttta ggctctagca atgttgaagc aagaatttct 88800
attccaatga actatatgaa caaggttaag attggagatg agtatgaatt tgatatcaga 88860
ggaatatcta caaaagctac gttagagaga ctggctccca tgtcaaccgg aggctccaat 88920
aataggtttag caatttttag atttgatacc ttttttaatc caggatcaat agtaaagctc 88980
aaactaagca tcaactgagaa agcaaaaggt acatgggttc caattaagtc actgtcccag 89040
tccgaacaag gtatttgggc catatatacc attaacgagc aacaagtagt tgtttagagat 89100
cttgttgatg ttattttatt tgaagacgaa tatgcttttg tcagcggaac acttaataat 89160
ggtgatttag taatttttag cgagactcaa aaaattattg aaggaaaaat aataaaataa 89220
aatgaatggtt attaattttc taatagaaaa gcctaggata ttatttctaa ctttagcatt 89280
tatattactt tctggaattt cttctgggct ttcagttcca attcaagaaa accctgaact 89340
ggctgagaga tgggggaggtg ttcgtatttt tcttcccggg gcatccccag aaagaattga 89400
aacagagata gtaaatgatc tagaaatcaa acttagagaa gttgaagaaa tcgatgagct 89460
tgaatcaatt attactcaag gtttttcaac aattgtagtt gaattaaatc aaagtgtacc 89520
tcctatactt attgaagaga cttggtccaa ggttcaagac aagctcaatc aaatagttat 89580
tcctcaaggt gcagaaatat ttcttgatag aagcagtggg ccgcctatca ctgttcagta 89640
tgctgtaacc tggaacggca gtggagatgc tccactaata atgatgtcca gactagcaag 89700

ccagctaaaa	agaaaattaa	gctcaatagg	ctcatctcat	caaactgcaa	tttttgggtga	89760
aacagatgaa	gaaatttttaa	ttgaactaga	ttcatcaaag	ctatcttcgc	ttggattatc	89820
atttcaagat	atcgcaagtg	ctattcaagc	cctagatgca	aaaaaaccta	ttgggtgtatc	89880
ctcaaacaac	aattctgagc	ttttatatag	actcaaagat	aatatacaga	gcattcaaaa	89940
actctcagaa	atacctatca	aggttattaa	taaatcagag	atcatacagc	tagatgatgt	90000
ggcatttatt	tcaaaaatcc	cggtttctcc	tattgaagac	atattcttgt	ttaatggaaa	90060
tgtagttatt	tctgttgctg	gaaccggatc	attttctcaa	agagtccatg	attatgtaga	90120
acgcgcaaca	attgttgtag	atgagatgag	agaaactctg	ccgactgaga	tcactataga	90180
tttagtttat	gacgaatctg	cttacacaac	taaaaaattt	aatgagcttg	taaaaagttt	90240
ttcattagca	atattttttg	tttttagcttt	aagtcttttt	tttcttgga	ttagatcagc	90300
aataattggt	actcttatcc	tgccattttc	tatttgcttg	gttatgattg	gttgtagggt	90360
tataggctta	cccttgcata	tgacatctat	cactggaatt	attattgcac	taggattgct	90420
tatagataat	gggattattg	ttgttgaaga	ttataagaat	agaagagcat	ctggattaa	90480
tatcaatgat	tcaatttcac	aaggactaaa	aaacttatgg	gctccattag	ctgctgcaac	90540
agcaacaacc	gtcttctctt	ttcttcttat	tgttactgga	gaaggggtcg	gtattgaatt	90600
cgtaggcggc	atggcaatga	cagtcattat	gtctataaca	tcttcattag	ttttggcggt	90660
tattaatgggt	ccagttctga	tgagttatat	ggaaaaaatt	ccgttcttta	aggatgtgga	90720
tattagcaag	gaaggggtata	gaaatgaaaa	aatccttaat	aaatataggg	ccttttttaa	90780
ctgggcgttc	ttagttccta	gaagagcaat	catgatatcg	cttgcatctg	ctgttctagg	90840
attctttctt	tttaattctt	tacctaaaga	tttctttcct	gctcaagata	gagatatgtt	90900
tagagttaat	atagaactgc	cttctaacgc	ctcatcactt	actacaatgc	agagagttaa	90960
ggaaattaga	gaagatatct	tagatagtga	tttaatttca	atagaaaaag	attattcggt	91020
tatcggcaga	atgatgccta	gagttttgat	gaatgttggt	ggtggagaag	aaaaacaagg	91080
atccaataat	attgcgcagt	ctgtattttt	tgctactgat	tattatgaaa	tgattgaaaa	91140

ccttccagat ttatcaagaa gactgggttaa aaataaccct gacattatag ttaatattga 91200
 tagtttctcg tctggccccc cggttttttc agatgttagt tatgtaattt ttggagatga 91260
 tccagattta ctaaaatcac ttggtgagga gctagagcta attattaaca attctcctga 91320
 tgtgagtctt acgaaatctg caacttcaaa ctcaataacc aatgttgagt ttgaacttaa 91380
 cagctcaaat atttcaactgt ctgggtcaaaa tgccaattat cttgtaaagt aaatgtttac 91440
 tgcaaacaat ggaatatttg ttggcactat gttggattca aacaaagaaa taccagtcag 91500
 gctgaaaggg ctgtctaata aaaacaatat tacgggaaat actagtttta taacaatgcc 91560
 ctctcaaggt ggttttgagt attttgatag tttcggaaaa agctcactaa caaacaatc 91620
 gtcaacaatt actaggcttg atggccaaag aacaaatgat gttgagggct ggatttggac 91680
 aggtacgctc ccgtctgcta ctgaaaaatc tattaataaa gatgttaaag attttgaatt 91740
 aagattgcca ataggctatt cattaaaaca acttggcgag gctgaaagca ggggccaatc 91800
 tcaagcctca ttatactctt cagcttttat gtatttcatt cttataatag taggcttggt 91860
 tatggcgctt aattctttca gagaggctgg tctaattttg tctgttgcatt tcttatcaat 91920
 ttggactatcc tttcttggtt tatttatagg ccagcaaaat tatggattta ttggaactat 91980
 aagtgcaata gggttaattg gcttatcaat aatgattca attattgtct tatctcacat 92040
 aaaagaagag gctgagaaga aatcactaac caaagctgag cttgttgaag ttgttatcag 92100
 gtctacacgt catataatca ctacctctt gacaacactt ggtggttttc ttccacttat 92160
 ttttgcaagt gtattcttca aaccgcttgc ttgggcaatg agcattggag tattaggcgc 92220
 gactattaca gccttattat atattcctgc aatgtttatg ataatgagaa aggttaagta 92280
 ctagaacaac tttccgagca ataatttctt cctagctttt ttacaataag tgattcatga 92340
 gtataggtct cgcaagaatc acatttaatc atagagttat caatcttttt tttatgaggt 92400
 cctgatatga gagatctgaa aaagaatacc gcagttatta tcaatataaa tatgaccaat 92460
 ggaatcaaaa gtaatatcga ttttaataag aacatttttt agtcactaga tgctggctta 92520
 ctatttctta taaaattcca agaccagcca gacttatctt cagatttttt tgattgctca 92580

taaagaaaat	attggaaact	actctgtgtt	ggatataaga	atagagacag	gaagaactca	94080
tcagattagg	gtccatctat	catcaaaaaa	actaccaata	attggagata	aaacctatga	94140
cccaagcagg	tctattgcaa	gagatacccc	tgaagagcta	attaatatta	tccgagggtt	94200
tccaaggcag	gcattacatg	caacacacct	ctcattcaat	gaccaaaaaa	caaataatat	94260
tttttctttt	gatattccca	ctccaaatga	tatggaggaa	ctacttctag	aattaagaaa	94320
attgatctaa	tagtaactaa	aaacttggtt	tttgattaat	aaaataatat	aaaccttatt	94380
cctaagagtt	ttttggtaag	aaattgaaat	tatctggcgc	agacatgcta	atgcaagcac	94440
ttcatgatga	aggtgttgag	ctaactcttg	gctaccagg	tggagccgcg	cttcatatct	94500
atgatgcaat	ttttagacaa	gataaaatag	atcatatttt	agtaaggcat	gagcaagggtg	94560
caacccatgc	agcagatgga	tattcaaggg	cgacaggtaa	gccaggagtt	gtcttagtca	94620
cttctggacc	tggtgcaaca	aatgctatta	caggaatcgc	gactgcattt	atggattcca	94680
taccaatgg	agttatttca	gggcagggtg	ctagccattt	aatagggtact	gatgcttttc	94740
aaagaaactga	tatgattgg	gtttcaagac	caattgttaa	gcatagctat	acagttttta	94800
atgctgaaga	aatacctaag	ataattaaag	aagcttttta	tgtcgcaact	tcaggcagac	94860
ctggacctgt	tgttatagat	atcccaaaag	acatgacagc	tccggataat	ctttttgatt	94920
actcgtatcc	tgaagaagcc	aagataagat	catacaatcc	tccgattgag	ccagaaaaaa	94980
atcaaataga	tagagcagtc	gaagctatat	tgatatcaaa	aaaaccagtt	atatatgctg	95040
gtggtggggc	aattgctagt	aatgccgaaa	aagaattact	tgaacttaat	gaaattattg	95100
atgctcctgt	tacaaatact	ttaatgggat	tgggtattta	ccctgctagt	catcatagat	95160
ttcttgggat	gttagggatg	catggaacat	atcaggcaaa	tatggcaatg	cataatgcag	95220
acttaataat	tgctattggc	gccagatttg	atgacaggat	taccaataaa	ccatcaaagt	95280
ttgcacctaa	tgccaaagtg	gttcatctag	atgttgatca	ctcatctgta	tcaaaaatta	95340
tagaagcaaa	tgtagctgtt	tttgggcaag	taaaaaatc	cttaaaatta	ataaaagaaa	95400
ctcttgaaaa	aaaattagac	tcttacgatt	ctttcgctct	tcagccttgg	cacgatcaga	95460

taaaagaatg gaaatcacta catgggttta attatgagct ttataaagat gaatctgatg	95520
atcatcccat tttaccccag gctgtagtcc agcatgtcca tgagattaca aatggggaag	95580
catatgtgac ttccgatgtt ggtcagcatc aaatgtttgc tgctcaatat tatcattttg	95640
ataagcctag aagatggatc aattctgggtg gtctaggaac tatggggtttt ggtttgccag	95700
cagcaatggg tgtaaaactc gctttttccaa aagatgaggt tgtttgcatt actggtgagg	95760
gtagtatcca aatgtgcatc caagaattgt ctacatgtct tcaatataat ctcccaataa	95820
aaataattaa tatcaataac gaagctcttg gtatgggttaa acaatggcag gatatgaatt	95880
atggaggaag gcactctgaa agtacctatc aaaactcgtt accggatttt ataaaactgg	95940
ctgaatcata tgggtcatata ggaattaaaa ttacaaaaaa ttctgattta agtgaaggct	96000
taaaaaaagc ttttgaaatg aaagataaac ttgtctttgt tgatatttat gtagatcctt	96060
cagagcatgt ttatccaatg cagggtgcaa atggcagcct agaaaatatg tggctatcaa	96120
aggatgaaca aacatgatta aaagaaaact aattttaatt atggaaaata aaccaggagc	96180
tctagtaaga gtagttggac tgtttcatca aagaggctac aacattgaaa cccttcatgt	96240
agatactgtt aaagactttt ctacttaca atcgatattg aaaaaaaact tgaaaccaa	96300
tttgaggata atcaaatac tagactgacc atagaaacaa tggtttcaga tgaccttatg	96360
aggcagatgt tgagacagct caataaatta atagatgtta tagctgtaag caatgaagag	96420
acaacctatt taaaaggagt attattagat gaaaatttat tatgacgaag atgcaaacat	96480
agaaattatt aaagggatga atgtctcaat aattggctat ggatctcaag ggaatgctca	96540
tgcaaataat cttcatgaat ctgggtgtaag tgttactgta ggtttaagag aagggtcttc	96600
ttcatgggca aaggcagaag aagcaggctt aaaagttcaa acagttgctg attcggtaat	96660
ccaagcagat ttggttatga ttttggcacc agatgaattc caaaaaaata tatatgaaac	96720
cgaaatcaag ccaaacttaa aaaccagtgc aattcttgca tttgcacatg gctttaatat	96780
tcattttgaa aaaatagttc ctgaagcaac taacagcgta attatgattg caccaaaagg	96840
tccaggccat actgttagaa gtacttatac caatgggtgga ggcgttccat ctctcatagc	96900

tggaatgcgt	agatattatg	aagaatacca	gtggcgaagg	cggcctcctg	gatctgtact	98400
gacactgagg	tgcgaaagcg	tgggtagcga	acaggattag	ataccctggg	agtccacgcc	98460
gtaaacgatg	acaactagct	gttgggagac	tatgtctttc	agtggcgag	ctaacgcttt	98520
aagttgtccg	cctggggagt	acggccgcaa	ggctaaaact	caaatgaatt	gacggggacc	98580
cgcacaagcg	gtggagcatg	tggtttaatt	cgatgcaacg	cgaaaaacct	tacctactct	98640
tgacatactt	ggaggctctt	gtaatgagag	tgtgcctttt	ggaaccaaga	tacaggtgct	98700
gcatggctgt	cgtcagctcg	tgtcgtgaga	tgttccgtta	agtcggataa	cgagcgcaac	98760
ccttaccctt	atgtgccagc	gattcgggtcg	ggaactataa	ggggactgcc	ggtgataaac	98820
cggaggaagg	tgaggacgac	gtcaagtcac	catggccctt	acgagtaggg	ctacacacgt	98880
gctacaatgg	gagatacaga	cggacgctaa	gccgcgaggt	ggtgctaata	ctaaaaagtc	98940
tttcgtagtc	cggattggag	tctgcaactc	gactccatga	agtcggaata	gctagtaata	99000
gcggatcagc	atgccgcggt	gaatacgttc	tcgggtcttg	tacacaccgc	ccgtcacacc	99060
atggaagtgg	attgcaccag	aagtagatag	tctaacctta	gggagggcgt	ttaccacggt	99120
gtgcttcatg	actgggggtga	agtcgtaaca	aggtagccgt	aggggaacct	gtggctggat	99180
cacctcetta	acgataaata	gcgtttttaa	cgccacacag	agtaatcaaa	tattaaaaaa	99240
aagaacatth	agatatgtaa	aatcattggt	atgtaatttt	ctagtgtata	catttatgta	99300
tacataagat	cactgcaatt	aaaaagtaac	atatgcattt	atgtgtatgt	taaaaaagta	99360
attaatatat	tttattaagt	tactctcaaa	aatgaagata	aaacttcaaa	aaaaatatgt	99420
aacctttttt	aagggtatat	gatcaagtaa	aggaagagca	caaggcggat	gccttggcag	99480
cataaggcga	tgaaggacgt	aataacctgc	gataagcctc	ggggagctgg	taaataagct	99540
tcgatccgag	gatttccgaa	tgggaaaacc	caatacacat	aagtgtatta	tcttatactg	99600
aatacatagg	tataagaggc	aaacctaggg	aactgaaaca	tctaagtacc	tagaggaaaa	99660
gaaatcaaca	gagattccgg	tagtagcggc	gagcgaaacc	ggaccagccc	ttaagcttat	99720
tttagtccag	caaaatatth	tggaaagttt	agccatagta	ggtgatagcc	ctgtatgcga	99780

ctctgcaaac tcgtaagagg aagtataggg tgtgacgcct gcccggtgcc ggaagggttaa 101280
 ttgatggggg tagcttatgc gaagctcttg atcgaagccc cggtaaacgg cggccgtaac 101340
 tataacgggc ctaaggtagc gaaattcctt gtcgggtaag ttccgacctg cacgaatggc 101400
 gtaacgatgg ccacactgtc tccaccatag actcagtga attgaaatcg ctgttaagat 101460
 gcagtgtacc cgcagctaga cggaaagacc ccgtgcacct ttactatagg ttgcgactgg 101520
 actttgacct tactttgtga ggataggtgg gagactttga agcagagacg ccagtctttg 101580
 tggagtcac cttgaaatac cacccttgta agattgaagt tctaacctag gtccattatc 101640
 tggatcaggg acagtgcgtg ctgggtagtt tgactggggc ggtctcctcc taaagagtaa 101700
 cggaggagta cgaaggatc cttatcacgg tcggacatcg tgaggtaagt ataaaggcag 101760
 aaggatgctt gactgcgaga tcgacggatc gagcaggtag gaaactaggt cttagtgatc 101820
 cgggtgggtt gaatggaagg gccatcgctc aacggataaa aggtacgccg gggataacag 101880
 gctgataccg cccaagagtt catatcgacg gcggtggttg gcacctcgat gtcggctcat 101940
 cacatcctgg ggctggagca ggtcccaagg gtatggctgt tcgccattta aagtggtagc 102000
 cgagctgggt ttagaacgtc gtgagacagt tcggtcccta tctgctgtgg gcgtttggag 102060
 atttgaggga agctgattct agtacgagag gaccgaattg gacgaacctc tgggtgttccg 102120
 gttgtcacgc cagtggcatt gccgggtagc tatgttcgga aaggataacc gctgaaagca 102180
 tataagcggg aagcctctcc caagattaaa tctcccagag actttatgtc tcctaaagag 102240
 tcgtcataga ctatgacgtt gataggcaag atgtgtaagc gctgcgaggc gttgagctaa 102300
 cttgtactaa taactcgtga ggcttgatca tgtaacctta agcaagggtc ataatttgag 102360
 taaaacattg tagtgagaat taaaaataa aaagttacat accagtttgc ctgatgacaa 102420
 tagcaacttg gaaccacctg atcccatctc gaactcagaa gtgaaacgag ttaacgcaa 102480
 tggtagtgca gggctctcct gtgtgagagt aggaaatcgt caggcttttt tctttaaggc 102540
 ttccagttta ctggaagcct ttttttttat ctcaagtata ataccagat gattattgga 102600
 ttaacaggag gcattgggtc tggcaaactc gccgctgcag acttctttat tgatttaggt 102660

attccttaaa gaatatatag cttgtctttg atcatttgat acatcgtcgt actctaaaag 104160
 atttttcctt gcatcaaagt ttctgctctc aattcttttt tgcgcatttt ctattcctct 104220
 tgaaagcatt ttagcttcaa tatgatcatc tcccatgccc agcctttcaa aagtagccct 104280
 tcttccatcc gaaataaaaa gtctcaagag atcgtcttct aaagataaga agaattttga 104340
 ataacctgga tcaccttgtc tgctgatct ccccttaac tgattatcta ttcttcttga 104400
 ttcatgcctt tctgttccaa gtatatgaag cccaccagat tctataactt ttttattatt 104460
 ttctttccac tctatatcgc ttgggtcttc tttttttccg ccaagaacta tatccgtccc 104520
 ccttcctgcc atattcgtag caatagttac cattccaggt ttacctgcat tggcaattat 104580
 ctctgcttct ttttcgtgat gcttggcatt taaaatttga tgaggatatt ttttgttatt 104640
 taaatatgct gatacttctt ctgaagattc aaccgaaact gttccaacaa gaatgggaga 104700
 agattttttt ctttaattggt caatttcttc aattagagct ttatatctcg attctgttgt 104760
 aagaaatact aagtcattaa gatcagctct agccatagga acatttggtg ggatgatgat 104820
 gacatttagg ccatagattt gactaaactc tactgcttca gtatctgctg ttctgtcat 104880
 cccagaaagt tttttaaata atctaaaaaa gttttggaat gtggtggatg ctagtggttg 104940
 agactctctt tggatagcaa cattttcttt gcattccagt gcctggtgaa caccttcact 105000
 cattcttctt ccgggcattg ttctacctgt atgctcatca atcaaaagaa cctcaccgtt 105060
 cctaaccaaa taatccacat tctttttaaa taagaagctt gctctaagtg ttgcttgaac 105120
 aaatttcata atttttaaat tagaaacaga gtaagcccat ctgaggctcc aagccgattc 105180
 cagc 105184

<210> 2

<211> 29

<212> DNA

<213> pcr primer

<220>

<221> oligo-nucleotide-primer

<222> (1)..(29)

<300>

<301> Beja,O., Aravind,L., Koonin,E.V., Suzuki,M.T., Hadd,A.,
Nguyen,L.P., Jovanovich,S.B., Gates,C.M., Feldman,R.A., Spudich,J.L.,
Spudich,E.N. and DeLong,E.F.

<302> Bacterial rhodopsin: evidence for a new type of phototrophy in
the sea

<303> Science

<304> 289

<305> 5486

<306> 1902-1906

<307> 2000-09-15

<309> _____

<400> 2

accatgggta aattattact gatattagg

29

<210> 3

<211> 24

<212> DNA

<213> pcr primer

<220>

<221> oligo-nucleotide-primer

<222> (1)..(24)

<300>

<301> Beja,O., Aravind,L., Koonin,E.V., Suzuki,M.T., Hadd,A.,
Nguyen,L.P., Jovanovich,S.B., Gates,C.M., Feldman,R.A., Spudich,J.L.,
Spudich,E.N. and DeLong,E.F.

<302> Bacterial rhodopsin: evidence for a new type of phototrophy in
the sea

<303> Science

<304> 289

<305> 5486

<306> 1902-1906

<307> 2000-09-15

<309> _____

<400> 3

agcattagaa gattctttaa cagc

24

<210> 4

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> light-driven proton pump; has the properties of a light-driven proton pump when expressed with retinal in Escherichia coli. Note that additional three nucleotide residues incorporated by pcr priming with reference to the original 31A08 DNA sequence (DNA residues 4-6, ggt), adding a new restriction site for cloning

<300>

<301> Beja,O., Aravind,L., Koonin,E.V., Suzuki,M.T., Hadd,A.,Nguyen,L.P., Jovanovich,S.B., Gates,C.M., Feldman,R.A.,Spudich,J.L., Spudich,E.N. and DeLong,E.F

<302> Bacterial rhodopsin: evidence for a new type of phototrophy in the sea

<303> Science

<304> 289

<305> 5486

<306> 1902-1906

<307> 2000-09-15

<308> AAG10475

<309> 2000-06-15

<313> (1)..(750)

<400> 4

atg	ggt	aaa	tta	tta	ctg	ata	tta	ggt	agt	ggt	att	gca	ctt	cct	aca	48
Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Val	Ile	Ala	Leu	Pro	Thr	
1			5					10					15			

ttt	gct	gca	ggt	ggt	ggt	gac	ctt	gat	gct	agt	gat	tac	act	ggt	ggt	96
Phe	Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val	
			20					25						30		

tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct gca act aat gtt gct gga tca tta ttt aag aaa tta	384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg	480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tac atg att tat gaa tta tgg gct gga gaa gga aaa tct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys	
165 170 175	
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat	576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr	
180 185 190	
att atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	
tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat	672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr	
210 215 220	

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 6

<211> 747

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(747)

<223> Native proteorhodopsin DNA sequence from BAC clone 31A08

<300>

<301> Beja,O., Aravind,L., Koonin,E.V., Suzuki,M.T., Hadd,A.,
Nguyen,L.P., Jovanovich,S.B., Gates,C.M., Feldman,R.A., Spudich,J.L.,
Spudich,E.N. and DeLong,E.F.

<302> Bacterial rhodopsin: evidence for a new type of phototrophy in
the sea

<303> Science

<304> 289

<305> 5486

<306> 1902-1906

<307> 2000-09-15

<309> _____

<313> (1)..(747)

<400> 6

atg	aaa	tta	tta	ctg	ata	tta	ggt	agt	ggt	att	gca	ctt	cct	aca	ttt	48
Met	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Val	Ile	Ala	Leu	Pro	Thr	Phe	
1				5					10					15		
gct	gca	ggt	ggt	ggt	gac	ctt	gat	gct	agt	gat	tac	act	ggt	ggt	tct	96
Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val	Ser	
			20					25					30			
ttt	tgg	tta	ggt	act	gct	gct	tta	tta	gca	tct	act	gta	ttt	ttc	ttt	144
Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe	Phe	
		35					40					45				

500 400 300 200 100

gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act gta Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr Val 50 55 60	192
tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg aga Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met Arg 65 70 75 80	240
ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac att Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr Ile 85 90 95	288
gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta att Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu Ile 100 105 110	336
ctt gct gct gca act aat gtt gct gga tca tta ttt aag aaa tta cta Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu Leu 115 120 125	384
gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca gga Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala Gly 130 135 140	432
atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg gta Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp Val 145 150 155 160	480
tac atg att tat gaa tta tgg gct gga gaa gga aaa tct gca tgt aat Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys Asn 165 170 175	528
act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat att Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr Ile 180 185 190	576
atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt tac Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly Tyr 195 200 205	624
ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat aac Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr Asn 210 215 220	672
ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg aat Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp Asn 225 230 235 240	720

gtt gct gtt aaa gaa tct tct aat gct
 Val Ala Val Lys Glu Ser Ser Asn Ala
 245

<210> 7

<211> 249

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 7

Met Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr Phe
 1 5 10 15

Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val Ser
 20 25 30

Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe Phe
 35 40 45

Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr Val
 50 55 60

Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met Arg
 65 70 75 80

Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr Ile
 85 90 95

Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu Ile
 100 105 110

Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu Leu
 115 120 125

Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala Gly
 130 135 140

Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp Val
 145 150 155 160

Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys Asn
 165 170 175

Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr Ile
 180 185 190

Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly Tyr
 195 200 205

Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr Asn
 210 215 220

Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp Asn
 225 230 235 240

Val Ala Val Lys Glu Ser Ser Asn Ala
 245

<210> 8

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> proteorhodopsin variant from clone EBAC40

<400> 8

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	

tct ttt tgg tta gtt act gct gct cta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	

gta tcg ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	

aga ggg gta tgg att gag act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	

att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	

att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta	384
Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu	
115 120 125	

ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gag gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	

gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg	480
Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp	
145 150 155 160	

gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt	528
---	-----

Val	Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Ala	Gly	Glu	Gly	Lys	Ala	Ala	Cys	
				165					170					175		
aat	act	gca	agt	cct	gct	gtg	caa	tca	gct	tac	aac	aca	atg	atg	tat	576
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Tyr	
			180					185					190			
ata	atc	atc	ttt	ggg	tgg	gca	att	tat	cct	gta	ggg	tat	ttc	aca	ggg	624
Ile	Ile	Ile	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly	
			195				200					205				
tac	cta	atg	ggg	gac	ggg	gga	tca	gct	ctt	aac	tta	aac	ctt	atc	tat	672
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr	
	210					215					220					
gac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggg	tta	att	ata	tgg	720
Asp	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp	
225					230					235					240	
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct							750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala							
				245					250							
<210>	9															
<211>	250															
<212>	PRT															
<213>	Naturally occurring gamma proteobacterium															
<400>	9															
Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Val	Ile	Ala	Leu	Pro	Thr	
1				5					10					15		
Phe	Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val	
			20					25					30			
Ser	Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe	
		35					40					45				

Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr	50	55	60	
Val	Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met	65	70	75	80
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr	85	90	95	
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu	100	105	110	
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Ala	Gly	Leu	Phe	Lys	Lys	Leu	115	120	125	
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala	130	135	140	
Gly	Ile	Met	Asn	Ala	Trp	Gly	Ala	Phe	Val	Ile	Gly	Cys	Leu	Ala	Trp	145	150	155	160
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Ala	Gly	Glu	Gly	Lys	Ala	Ala	Cys	165	170	175	
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Tyr	180	185	190	
Ile	Ile	Ile	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly	195	200	205	
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr	210	215	220	
Asp	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp	225	230	235	240

att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct gct act aat gtt gct gga tca tta ttt aag aaa tta	384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg	480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tac atg att tat gaa cta tgg gct gga gaa gga aaa tct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys	
165 170 175	
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat	576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr	
180 185 190	
att atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	
tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat	672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr	
210 215 220	
aac ctt gct gat ttt gtt aac aag att cta ttt ggt tta att ata tgg	720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp	
225 230 235 240	
aat gtt gct gtt aaa gaa tct tct aat gct	750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala	
245 250	

<210> 11

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 11

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Ala Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 12

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from clone EBAC64

<400> 12

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt	96
---	----

Phe	Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val	
			20					25					30			
tct	ttt	tgg	tta	gtt	aca	gct	gct	cta	tta	gca	tct	act	gta	ttt	ttc	144
Ser	Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe	
		35					40					45				
ttt	gtt	gaa	aga	gat	aga	gtt	tct	gca	aaa	tgg	aaa	aca	tca	tta	act	192
Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr	
	50					55					60					
gta	tct	ggt	ctt	gtt	act	ggt	att	gct	ttc	tgg	cat	tac	atg	tac	atg	240
Val	Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met	
65					70					75					80	
aga	gga	gta	tgg	att	gaa	act	ggt	gat	tcg	cct	act	gta	ttt	aga	tac	288
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr	
				85					90					95		
att	gat	tgg	tta	cta	aca	gtt	cct	tta	tta	ata	tgt	gaa	ttc	tac	tta	336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu	
			100					105					110			
att	ctt	gct	gct	gca	act	aat	gtt	gcc	ggc	tca	tta	ttt	aag	aaa	ctt	384
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Gly	Ser	Leu	Phe	Lys	Lys	Leu	
		115					120					125				
cta	gtt	ggt	tct	ctt	gtt	atg	ctt	gtg	ttt	ggt	tac	atg	ggt	gaa	gca	432
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala	
	130					135					140					
gga	att	atg	gca	gct	tgg	cct	gca	ttc	att	att	ggg	tgt	tta	gct	tgg	480
Gly	Ile	Met	Ala	Ala	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Cys	Leu	Ala	Trp	
145					150					155					160	
gta	tac	atg	att	tat	gaa	cta	tat	gct	gga	gaa	gga	aaa	tct	gca	tgt	528
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys	
				165					170					175		
aat	act	gca	agt	cct	tcg	gtt	caa	tca	gct	tac	aac	aca	atg	atg	gct	576
Asn	Thr	Ala	Ser	Pro	Ser	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Ala	
			180					185					190			
atc	ata	gtc	ttc	ggt	tgg	gca	att	tat	cct	ata	ggt	tat	ttc	aca	ggt	624
Ile	Ile	Val	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Ile	Gly	Tyr	Phe	Thr	Gly	
		195					200					205				
tac	cta	atg	ggt	gac	ggt	gga	tca	gct	ctt	aac	tta	aac	ctt	att	tat	672

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 13

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 13

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Ile Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 14

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone HOT01m: GenBank# AF349978

<400> 14

atg	ggt	aaa	tta	tta	ctg	ata	tta	ggt	agt	ggt	att	gca	ctt	cct	aca	48
Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Val	Ile	Ala	Leu	Pro	Thr	
1			5					10					15			

ttt	gct	gca	ggt	ggt	ggt	gac	ctt	gat	gct	agt	gat	tac	act	ggt	ggt	96
Phe	Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val	
			20					25					30			

tct	ttt	tgg	tta	ggt	act	gct	gct	cta	tta	gca	tct	act	gta	ttt	ttc	144
Ser	Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe	
		35					40					45				

ttt	ggt	gaa	aga	gat	aga	ggt	tct	gca	aaa	tgg	aaa	aca	tca	tta	act	192
Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr	
	50					55					60					

gta	tcg	ggt	ctt	ggt	act	ggt	att	gct	ttc	tgg	cat	tac	atg	tac	atg	240
Val	Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met	
65					70					75					80	

aga	ggg	gta	tgg	att	gag	acc	ggt	gat	tcg	cca	act	gta	ttt	aga	tac	288
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr	
				85				90						95		

att	gat	tgg	tta	cta	aca	ggt	cct	cta	ttg	ata	tgt	gaa	ttc	tac	tta	336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu	
			100					105					110			

att	ctt	gct	gct	gca	aca	aat	ggt	gct	gct	ggc	ctg	ttt	aag	aaa	tta	384
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Ala	Gly	Leu	Phe	Lys	Lys	Leu	
		115					120					125				

ttg	ggt	ggt	tct	ctt	ggt	atg	ctt	gtg	ttt	ggt	tac	atg	ggt	gag	gca	432
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala	

130	135	140	
gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg			480
Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp			
145	150	155	160
gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt			528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys			
	165	170	175
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat			576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr			
	180	185	190
ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt			624
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly			
	195	200	205
tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat			672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr			
	210	215	220
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg			720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp			
	225	230	235
aat gtt gct gtt aaa gaa tct tct aat gct			750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala			
	245	250	

<210> 15

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 15

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val

20

25

30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr

210

215

220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 16

<211> 753

<212> DNA

<213> Naturally occurring gamma prtoeobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone HOT75m1: GenBank#AF349979

<400> 16

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt gct 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Ala
 50 55 60

gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met 65 70 75 80	240
aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr 85 90 95	288
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu 100 105 110	336
att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu 115 120 125	384
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala 130 135 140	432
gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp 145 150 155 160	480
tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val 165 170 175	528
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met 180 185 190	576
att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly 195 200 205	624
tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile 210 215 220	672
tat aac ctt gcc gac ctt gtt aac aag att cta ttt ggt ttg atc att Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile 225 230 235 240	720
tgg aat gtt gct gtt aaa gaa tct tct aat gct Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala 245 250	753

<210> 17

<211> 251

<212> PRT

<213> Naturally occurring gamma prtoeobacterium

<400> 17

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Ala
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 18

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone HOT75m3; GenBank#AF349980

<400>	18															
atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca																48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser																
1	5 10 15															
ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt																96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val																
	20 25 30															
tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gta ttc ttt																144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe																
	35 40 45															
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act																192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr																
	50 55 60															
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg																240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met																
65	70 75 80															
aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat																288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr																
	85 90 95															
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta																336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu																
	100 105 110															
att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt																384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu																
	115 120 125															
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct																432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala																
	130 135 140															
ggt tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg																480
Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp																
145	150 155 160															
tta tac atg att tat gag cta cat atg ggt gaa ggt aag gct gct gta																528
Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val																
	165 170 175															

agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg aag 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys
 180 185 190

att att gtt att gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

tac cta atg agt ggt gac ggt gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 19

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 19

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val	Ser	Gly	Leu	Ile	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Leu	Tyr	Met	65	70	75	80
Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr	85	90	95	
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu	100	105	110	
Ile	Leu	Ala	Ala	Cys	Thr	Ser	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu	115	120	125	
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Phe	Ala	Gly	Glu	Ala	130	135	140	
Gly	Leu	Ala	Pro	Val	Leu	Pro	Ala	Phe	Ile	Ile	Gly	Met	Ala	Gly	Trp	145	150	155	160
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	His	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val	165	170	175	
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Lys	180	185	190	
Ile	Ile	Val	Ile	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly	195	200	205	
Tyr	Leu	Met	Ser	Gly	Asp	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile	210	215	220	
Tyr	Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	225	230	235	240
Trp	Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala						245	250		

<210> 20

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone HOT75m4; GenBank #AF349981

<400> 20

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	
ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	
tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	
aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat	288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta tta act gtt cca tta caa gtg gtt gag ttc tat cta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu	

100	105	110	
att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt			384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu			
115	120	125	
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct			432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala			
130	135	140	
gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg			480
Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp			
145	150	155	160
tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta			528
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val			
165	170	175	
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg			576
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met			
180	185	190	
att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt			624
Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly			
195	200	205	
tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata			672
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile			
210	215	220	
tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att			720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile			
225	230	235	240
tgg aat gtt gct gtt aaa gaa tct tct aat gct			753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala			
245	250		

<210> 21

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 21

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 22

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone HOT75m8: GenBank#AF349982

<400> 22

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val

20	25	30	
tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe 35 40 45			144
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr 50 55 60			192
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met 65 70 75 80			240
aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr 85 90 95			288
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu 100 105 110			336
att ctt gct gct tgt aca aat gtt gct gct tca tta ttt aag aag ctt Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu 115 120 125			384
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala 130 135 140			432
gga ttg gct cct gta tgg cct gct ttc att att ggt atg gct gga tgg Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp 145 150 155 160			480
tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val 165 170 175			528
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg gtg Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val 180 185 190			576
att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly 195 200 205			624
tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile			672

210		215		220	
tat aac ctt gcc gac ctt gtt aac aag att cta ttt ggt ttg atc att					720
Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile					
225		230		235	240

tgg aat gtt gct gtt aaa gaa tct tct aat gct					753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala					
	245			250	

<210> 23

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 23

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu

100
 105
 110
 115
 120
 125
 130
 135
 140
 145
 150
 155
 160
 165
 170
 175
 180
 185
 190
 195
 200
 205
 210
 215
 220
 225
 230
 235
 240
 245
 250

Ile	Leu	Ala	Ala	Cys	Thr	Asn	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu
		115					120					125			
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Phe	Ala	Gly	Glu	Ala
	130					135					140				
Gly	Leu	Ala	Pro	Val	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Met	Ala	Gly	Trp
145					150					155					160
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val
				165					170					175	
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Val
			180					185					190		
Ile	Ile	Val	Val	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly
		195					200					205			
Tyr	Leu	Met	Gly	Gly	Glu	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile
	210					215					220				
Tyr	Asn	Leu	Ala	Asp	Leu	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile
225					230					235					240
Trp	Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala					
				245					250						

<210> 24

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB0m1: GenBank#AF349983

<400> 24

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

tct ttt tgg tta gtt act gct gct cta tta gca tct act gta ttt ttc 144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

tga ggg gta tgg att gag act ggt gat tcg cca act gta ttt aga tac 288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta 336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta 384
Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
115 120 125

ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gag gca 432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

gga att atg aac gct tgg cct gca ttc att att ggg tgt tta gct tgg 480
 Gly Ile Met Asn Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160

gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt 528
 Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct 576
 Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
 180 185 190

atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct 750
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 25

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 25

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Asn Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Val 65	Ser	Gly	Leu	Val	Thr 70	Gly	Ile	Ala	Phe	Trp 75	His	Tyr	Met	Tyr	Met 80	
aga	ggg	gta	tgg	att	gaa	act	ggg	gat	tcg	cca	act	gta	ttt	aga	tac	288
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr	
				85					90					95		
att	gat	tgg	tta	cta	aca	ggt	cct	cta	tta	ata	tgt	gaa	ttc	tac	tta	336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu	
			100					105					110			
att	ctt	gct	gct	gct	act	aat	ggt	gct	gct	ggc	ctg	ttt	aag	aaa	tta	384
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Ala	Gly	Leu	Phe	Lys	Lys	Leu	
		115					120					125				
ttg	gtt	ggg	tct	ctt	gtt	atg	ctt	gtg	ttt	ggg	tac	atg	ggg	gaa	gca	432
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala	
	130					135					140					
gga	att	atg	aac	gct	tgg	ggg	gca	ttc	gtt	att	ggg	tgt	tta	gct	tgg	480
Gly	Ile	Met	Asn	Ala	Trp	Gly	Ala	Phe	Val	Ile	Gly	Cys	Leu	Ala	Trp	
145					150					155					160	
gta	tac	atg	att	tat	gag	ctt	tgg	ctt	gga	gaa	gga	aaa	gct	gcg	tgt	528
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Leu	Gly	Glu	Gly	Lys	Ala	Ala	Cys	
				165					170					175		
aat	aca	gca	agt	cct	gct	gtt	cag	tca	gct	tac	aac	aca	atg	atg	atg	576
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Met	
			180				185						190			
atc	atc	atc	ttt	ggg	tgg	gca	att	tat	cct	gta	ggg	tat	ttc	aca	ggg	624
Ile	Ile	Ile	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly	
		195				200						205				
tac	cta	atg	ggg	gac	ggg	gga	tca	gca	ctt	aac	tta	aac	ctt	atc	tat	672
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr	
	210					215					220					
aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggg	tta	att	ata	tgg	720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp	
225					230					235					240	
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct							750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala							
				245					250							

<210> 27

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 27

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Leu Gly Glu Gly Lys Ala Ala Cys
 165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Met
 180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 28

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB20m2; GenBank #AF349985

Table 1. Demographic characteristics of the study population	
Age (years)	50.0 ± 10.0
Gender	
Male	50.0%
Female	50.0%
Education (years)	12.0 ± 2.0
Occupation	
Professional	30.0%
Managerial	20.0%
Technical	10.0%
Service	20.0%
Unemployed	20.0%
Marital status	
Married	70.0%
Single	10.0%
Divorced	10.0%
Widowed	10.0%
Health status	
Good	80.0%
Fair	10.0%
Poor	10.0%
Smoking status	
Smoker	30.0%
Non-smoker	70.0%
Alcohol consumption	
Regular	10.0%
Occasional	20.0%
Never	70.0%
Family size	3.0 ± 1.0
Income (USD/month)	1000.0 ± 500.0
Health insurance	
Yes	90.0%
No	10.0%

Page 120

Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Tyr	
			180					185					190			
ata	atc	atc	ttt	ggt	tgg	gca	att	tat	cct	gta	ggt	tat	ttc	aca	ggt	624
Ile	Ile	Ile	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly	
			195				200					205				
tac	cta	atg	ggt	gac	ggt	gga	tca	gct	ctt	aac	tta	aac	ctt	atc	tat	672
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr	
	210					215					220					
aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg	720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp	
225					230					235					240	
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct							750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala							
				245					250							

<210> 29

<211> 250

<212> PRT

<213> Naturally occuring gamma proteobacterium

<400> 29

Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Val	Ile	Ala	Leu	Pro	Thr	
1				5					10					15		
Phe	Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val	
			20					25					30			
Ser	Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe	
		35					40					45				
Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr	
	50					55					60					

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

```

<210> 30
<211> 750
<212> DNA
<213> Naturally occurring gamma proteobacterium

<220>
<221> CDS
<222> (1)..(750)
<223> Proteorhodopsin variant from pcr clone MB20m5; GenBank#AF349986

<400> 30
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca      48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1          5          10          15

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt      96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
          20          25          30

tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc      144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
          35          40          45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act      192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
          50          55          60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg      240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65          70          75          80

aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac      288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
          85          90          95

att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta      336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
          100          105          110

```


<400> 31

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gln Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Leu Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 32

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB20m12; GenBank #AF349987

<400> 32

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	

tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc	144
---	-----

Ser	Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe	
		35					40					45				
ttt	ggt	gaa	aga	gat	aga	ggt	tct	gca	aaa	tgg	aaa	aca	tca	tta	act	192
Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr	
	50					55				60						
gta	tct	ggt	ctt	ggt	act	ggt	att	gct	ttc	tgg	cat	tac	atg	tac	atg	240
Val	Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met	
65					70					75					80	
aga	ggg	gta	tgg	att	gaa	act	ggt	gat	tcg	cca	act	gta	ttt	aga	tac	288
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr	
				85				90					95			
att	gat	tgg	tta	cta	aca	ggt	cct	cta	tta	ata	tgt	gaa	ttc	tac	tta	336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu	
			100					105					110			
att	ctt	gct	gct	gca	gct	aat	ggt	gct	gga	tca	tta	ttt	aag	aaa	tta	384
Ile	Leu	Ala	Ala	Ala	Ala	Asn	Val	Ala	Gly	Ser	Leu	Phe	Lys	Lys	Leu	
		115					120					125				
cta	ggt	ggt	tct	ctt	ggt	atg	ctt	gtg	ttt	ggt	tac	atg	ggt	gaa	gca	432
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala	
	130					135					140					
gga	atc	atg	gct	gca	tgg	cct	gca	ttc	att	att	ggg	tgt	tta	gct	tgg	480
Gly	Ile	Met	Ala	Ala	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Cys	Leu	Ala	Trp	
145					150					155					160	
gta	tac	atg	att	tat	gaa	tta	tgg	gct	gga	gaa	gga	aaa	tct	gca	tgt	528
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys	
				165				170						175		
aat	act	gca	agt	cct	gct	gtg	caa	tca	gcc	tac	aac	aca	atg	atg	tat	576
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Tyr	
			180					185					190			
att	atc	atc	ttt	ggt	tgg	gcg	att	tat	cct	gta	ggt	tat	ttc	aca	ggt	624
Ile	Ile	Ile	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly	
		195					200					205				
tac	ttg	atg	ggt	gac	ggt	gga	tca	gct	ctt	aac	tta	aac	ctt	atc	tat	672
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr	
	210					215					220					
aac	ctt	gct	gac	ttt	ggt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg	720

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

750

<210> 33

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 33

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Ala Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 34

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB40m1; GenBank #AF349988

<400> 34

atg ggt aaa tta tta ctg ata ata ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Ile Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	
ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	
tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt	384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg	480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp	

145					150					155				160			
gta	tat	atg	att	tat	gaa	cta	tat	gct	gga	gaa	gga	aaa	tct	gca	tgt		528
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys		
				165					170					175			
aat	aca	gca	agt	cct	gct	gtg	caa	tca	gct	tac	aac	aca	atg	atg	tat		576
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Tyr		
			180					185					190				
att	atc	gtc	ttt	ggc	tgg	gcg	att	tat	cct	gta	ggc	tat	ttc	aca	ggc		624
Ile	Ile	Val	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly		
		195					200					205					
tac	ctg	atg	ggc	gac	ggc	gga	tca	gct	ctt	aac	tta	aac	ctt	atc	tat		672
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr		
	210					215					220						
aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggc	tta	att	ata	tgg		720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp		
225					230					235					240		
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct								750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala								
				245					250								

<210> 35

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 35

Met Gly Lys Leu Leu Leu Ile Ile Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe

FOOTNOTES

35

40

45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp

aga ggg gta tgg att gag act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta	384
Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu	
115 120 125	
ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gag gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg	480
Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys	
165 170 175	
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat	576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr	
180 185 190	
ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	
tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat	672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr	
210 215 220	
aac ctt gct gac ttt gtt aac aag aat cta ttt ggt tta att ata tgg	720
Asn Leu Ala Asp Phe Val Asn Lys Asn Leu Phe Gly Leu Ile Ile Trp	
225 230 235 240	
aat gtt gct gtt aaa gaa tct tct aat gct	750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala	
245 250	

<210> 37

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 37

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Asn Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 38

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB40m12; GenBank # AF34999

<400> 38

atg ggt aaa tta tta cgg ata tta ggt agt gtt att gca ctt cct aca Met Gly Lys Leu Leu Arg Ile Leu Gly Ser Val Ile Ala Leu Pro Thr 1 5 10 15	48
ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val 20 25 30	96
tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe 35 40 45	144
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr 50 55 60	192
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met 65 70 75 80	240
aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr 85 90 95	288
att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu 100 105 110	336
att ctt gct gct gca act aat gtt gct gga tca tta ttt aag aaa tta Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu 115 120 125	384
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala 130 135 140	432
gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp 145 150 155 160	480
gta tac atg att tat gaa cta tgg gct gga gaa gga aaa tct gca tgt Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys 165 170 175	528
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr 180 185 190	576

atc atc atc gtt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	

tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat	672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr	
210 215 220	

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg	720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp	
225 230 235 240	

aat gtt gct gtt aaa gaa tct tct aat gct	750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala	
245 250	

<210> 39

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 39

Met Gly Lys Leu Leu Arg Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB100m5; GenBank #AF349991

<400> 40

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	
ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	
tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt	384

Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Gly	Ser	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
cta	gtt	ggt	tct	ctt	gtt	atg	ctt	gtg	ttt	ggt	tac	atg	ggt	gaa	gca		432
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala		
	130					135					140						
gga	att	atg	gca	gct	tgg	cct	gca	ttc	att	att	ggg	tgt	tta	gct	tgg		480
Gly	Ile	Met	Ala	Ala	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Cys	Leu	Ala	Trp		
145					150					155					160		
gta	tac	atg	att	tat	gaa	cta	tat	gct	gga	gaa	gga	aaa	tct	gca	tgt		528
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys		
				165					170					175			
aat	act	gca	agt	cct	tcg	gtt	caa	tca	gct	tac	aac	aca	atg	atg	gct		576
Asn	Thr	Ala	Ser	Pro	Ser	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Ala		
			180					185					190				
atc	ata	gtc	ttc	ggg	tgg	gca	att	tat	cct	gta	ggt	tat	ttc	aca	ggt		624
Ile	Ile	Val	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly		
		195					200					205					
tac	cta	atg	ggg	gac	ggg	gga	tca	gct	ctt	aac	tta	aac	ctt	att	tat		672
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr		
	210					215					220						
aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg		720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp		
225					230					235					240		
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct								750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala								
				245					250								

<210> 41

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 41

Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Val	Ile	Ala	Leu	Pro	Thr	
1				5					10					15		
Phe	Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val	
			20					25					30			
Ser	Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe	
		35					40					45				
Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr	
	50					55					60					
Val	Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met	
65					70					75					80	
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr	
				85					90					95		
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu	
			100					105					110			
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Gly	Ser	Leu	Phe	Lys	Lys	Leu	
		115					120					125				
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala	
		130				135					140					
Gly	Ile	Met	Ala	Ala	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Cys	Leu	Ala	Trp	
145					150					155					160	
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys	
				165					170					175		
Asn	Thr	Ala	Ser	Pro	Ser	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Ala	
			180					185					190			

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 42

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB100m7; GenBank #AF349992

<400> 42

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct gct act aat gtt gcc ggc tca tta ttt aag aaa ctt	384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg	480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys	
165 170 175	
aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct	576
Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala	
180 185 190	
atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	
tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat	672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr	
210 215 220	
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg	720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp	
225 230 235 240	

aat gct gct gtt aaa gaa tct tct aat gct
 Asn Ala Ala Val Lys Glu Ser Ser Asn Ala
 245 250

750

<210> 43

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 43

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Ala Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 44

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB100m9; GenBank #AF349993

<400> 44

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg 240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

ata gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta 336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

att ctt gcc gct gca act aat gtt gct gga tca tta ttt aag aaa tta 384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

ctt gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg 480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

gta tac atg att tat gaa cta tgg gct gga gaa gga aaa tct gca tgt 528

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

atc atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

tac ctt atg ggt gac ggt gga tca gca ctt aac tta aac ctt att tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct 750
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 45

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 45

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 46

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB100m10; GenBank #AF34999

<400> 46

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	

tct ttt tgg tta gtt aca gct gct cta tta gcg tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	

aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	

85										90					95					
att	gat	tgg	tta	cta	aca	gtt	cct	tta	tta	ata	tgt	gaa	ttc	tac	tta		336			
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu					
			100					105					110							
att	ctt	gct	gct	gca	act	aat	gtt	gcc	ggc	tca	tta	ttt	aag	aaa	ctt		384			
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Gly	Ser	Leu	Phe	Lys	Lys	Leu					
		115					120					125								
cta	gtt	ggg	tct	ctt	gtt	atg	ctt	gtg	ttt	ggg	tac	atg	ggg	gaa	gca		432			
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala					
	130					135					140									
gga	ata	atg	gcg	gct	tgg	cct	gca	ttc	atc	gtt	gga	tgt	tta	gca	tgg		480			
Gly	Ile	Met	Ala	Ala	Trp	Pro	Ala	Phe	Ile	Val	Gly	Cys	Leu	Ala	Trp					
145					150					155					160					
gta	tat	atg	att	tat	gaa	cta	tgg	gct	ggg	gaa	gga	aaa	tct	gca	tgt		528			
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys					
				165					170					175						
aat	act	gca	agt	cct	gct	gta	cag	tca	gct	tac	aac	aca	atg	atg	tat		576			
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Tyr					
			180					185					190							
atc	atc	atc	gtt	ggg	tgg	gca	att	tat	cct	gta	ggg	tat	ttc	aca	ggg		624			
Ile	Ile	Ile	Val	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly					
		195				200						205								
tac	cta	atg	ggg	gac	ggg	gga	tca	gct	ctt	aat	cta	aac	ctt	att	tat		672			
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr					
	210					215					220									
aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggg	tta	att	ata	tgg		720			
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp					
225					230					235					240					
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct								750			
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala											
				245				250												

<210> 47

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 47

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Val Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 48

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PALB1; GenBank #AF349995

<400> 48

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

ttt	gct	gct	gct	ggt	ggc	gat	cta	gat	ata	agt	gat	act	gtt	ggt	ggt		96
Phe	Ala	Ala	Ala	Gly	Gly	Asp	Leu	Asp	Ile	Ser	Asp	Thr	Val	Gly	Val		
			20					25					30				
tca	ttc	tgg	ctg	gtt	aca	gct	ggt	atg	tta	gcg	gca	act	gta	ttc	ttt		144
Ser	Phe	Trp	Leu	Val	Thr	Ala	Gly	Met	Leu	Ala	Ala	Thr	Val	Phe	Phe		
		35					40					45					
ttt	gta	gaa	aga	gac	caa	gtc	agc	gct	aag	tgg	aaa	act	tca	ctt	act		192
Phe	Val	Glu	Arg	Asp	Gln	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr		
	50					55					60						
gta	tct	ggt	tta	att	act	ggt	ata	gct	ttt	tgg	cat	tat	ctc	tac	atg		240
Val	Ser	Gly	Leu	Ile	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Leu	Tyr	Met		
65					70				75						80		
aga	ggt	gtt	tgg	ata	gat	act	ggt	gat	aca	cca	aca	gta	ttt	aga	tat		288
Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr		
				85					90					95			
att	gat	tgg	cta	tta	act	gtt	cca	tta	caa	atg	gtt	gag	ttc	tat	cta		336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gct	gct	tgt	aca	agt	gtt	gct	gct	tca	tta	ttt	aag	aag	ctt		384
Ile	Leu	Ala	Ala	Cys	Thr	Ser	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
cta	gct	ggt	tca	tta	gta	atg	tta	ggt	gct	gga	ttt	gca	ggc	gaa	gct		432
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Phe	Ala	Gly	Glu	Ala		
	130					135					140						
ggt	tta	gct	cct	gta	tta	cct	gct	ttc	att	ctt	ggt	atg	gct	ggt	tgg		480
Gly	Leu	Ala	Pro	Val	Leu	Pro	Ala	Phe	Ile	Leu	Gly	Met	Ala	Gly	Trp		
145					150					155					160		
tta	tac	atg	att	tat	gag	cta	cat	atg	ggt	gaa	ggt	aag	gct	gct	gta		528
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	His	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val		
				165				170						175			
agt	act	gca	agt	cct	gct	gtt	aac	tct	gct	tac	aat	gca	atg	atg	aag		576
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Lys		
			180					185					190				
att	att	gtt	att	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggt		624
Ile	Ile	Val	Ile	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly		
		195					200					205					

tac cta atg agt ggt gac ggt gta tac gct tca aac tta aac ctt ata 672
Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 49

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 49

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Leu Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys
180 185 190

Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 50
<211> 753
<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PALB2; GenBank #AF349996

<400> 50

atg	ggt	aaa	tta	tta	ctg	ata	tta	ggt	agt	gct	att	gca	ctt	cca	tca	48
Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Ala	Ile	Ala	Leu	Pro	Ser	
1			5					10					15			

ttt	gct	gct	gct	ggt	ggc	gat	cta	gat	ata	agt	gat	act	ggt	ggt	ggt	96
Phe	Ala	Ala	Ala	Gly	Gly	Asp	Leu	Asp	Ile	Ser	Asp	Thr	Val	Gly	Val	
			20					25					30			

tca	ttc	tgg	ctg	gtt	aca	gct	ggt	atg	tta	gcg	gca	act	gtg	ttc	ttt	144
Ser	Phe	Trp	Leu	Val	Thr	Ala	Gly	Met	Leu	Ala	Ala	Thr	Val	Phe	Phe	
		35					40					45				

ttt	gta	gaa	aga	gac	caa	gtc	agc	gct	gag	tgg	aaa	act	tca	ctt	act	192
Phe	Val	Glu	Arg	Asp	Gln	Val	Ser	Ala	Glu	Trp	Lys	Thr	Ser	Leu	Thr	
	50					55					60					

gta	tct	ggt	tta	att	act	ggt	ata	gct	ttt	tgg	cat	tat	ctc	tat	atg	240
Val	Ser	Gly	Leu	Ile	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Leu	Tyr	Met	
65					70				75					80		

aga	ggt	gtt	tgg	ata	gat	act	ggt	gat	acc	cca	aca	gta	ttc	aga	tat	288
Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr	
			85					90					95			

att	gat	tgg	tta	tta	act	gtt	cca	tta	caa	atg	gtt	gag	ttc	tat	cta	336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu	
			100					105					110			

att	ctt	gct	gct	tgt	aca	agt	gtt	gct	gct	tca	tta	ttt	aag	aag	ctt	384
Ile	Leu	Ala	Ala	Cys	Thr	Ser	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu	
		115					120					125				

cta	gct	ggt	tca	tta	gta	atg	tta	ggt	gct	gga	ttt	gca	ggc	gaa	gct	432
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 52

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PALB5; GenBank#AF349997

<400> 52

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	

gta Val 65	tct Ser	ggt Gly	tta Leu	att Ile	act Thr 70	ggt Gly	ata Ile	gcc Ala	ttt Phe	tgg Trp 75	cat His	tat Tyr	ctc Leu	tat Tyr	atg Met 80	240
aga Arg	ggt Gly	gtt Val	tgg Trp	ata Ile 85	gac Asp	act Thr	ggt Gly	gat Asp	acc Thr 90	cca Pro	aca Thr	gta Val	ttc Phe	aga Arg 95	tat Tyr	288
att Ile	gat Asp	tgg Trp	tta Leu 100	tta Leu	act Thr	ggt Val	cca Pro	tta Leu 105	caa Gln	atg Met	ggt Val	gag Glu	ttc Phe 110	tat Tyr	cta Leu	336
att Ile	ctt Leu	gct Ala 115	gct Ala	tgt Cys	aca Thr	aat Asn	ggt Val 120	gct Ala	gct Ala	tca Ser	tta Leu	ttt Phe 125	aag Lys	aag Lys	ctt Leu	384
cta Leu	gct Ala 130	ggt Gly	tca Ser	tta Leu	gta Val	atg Met 135	tta Leu	ggt Gly	gct Ala	gga Gly	ttt Phe 140	gca Ala	ggc Gly	gaa Glu	gct Ala	432
gga Gly 145	tta Leu	gct Ala	cct Pro	gta Val	tgg Trp 150	cct Pro	gct Ala	ttc Phe	att Ile	att Ile 155	ggt Gly	atg Met	gct Ala	gga Gly	tgg Trp 160	480
tta Leu	tac Tyr	atg Met	att Ile	tat Tyr 165	gag Glu	cta Leu	tat Tyr	atg Met	ggt Gly 170	gaa Glu	ggt Gly	aag Lys	gct Ala	gct Ala 175	gta Val	528
agt Ser	act Thr	gca Ala	agt Ser 180	cct Pro	gct Ala	ggt Val	aac Asn	tct Ser	gca Ala	tac Tyr	aac Asn	gca Ala	atg Met 190	atg Met	atg Met	576
att Ile	att Ile	gtt Val 195	gtt Val	gga Gly	tgg Trp	gca Ala	att Ile 200	tat Tyr	cct Pro	gct Ala	gga Gly	tat Tyr 205	gct Ala	gct Ala	ggt Gly	624
tac Tyr	cta Leu 210	atg Met	ggt Gly	ggc Gly	gaa Glu	ggt Gly 215	gta Val	tac Tyr	gct Ala	tca Ser	aac Asn 220	cta Leu	aac Asn	ctt Leu	ata Ile	672
tat Tyr 225	aac Asn	ctt Leu	gct Ala	gac Asp	ttt Phe 230	ggt Val	aac Asn	aag Lys	att Ile	cta Leu 235	ttt Phe	ggt Gly	ttg Leu	atc Ile	att Ile 240	720
tgg Trp	aat Asn	gtt Val	gct Ala	ggt Val 245	aaa Lys	gaa Glu	tct Ser	tct Ser	aat Asn 250	gct Ala						753

<210> 53

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 53

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 54

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalB7; GenBank #AF349999

<400> 54

atg ggt aaa tta tta ctg ata tta ggt agt gct att gcg ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	
ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	
tca ttc tgg ctg gtt acg gct ggt atg tta gcg gca act gta ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	
aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat	288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu	
100 105 110	
att ctt gcc gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt	384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gct ggt tca ttg gta atg tta ggt gct gga tct gca ggc gaa gct	432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Ser Ala Gly Glu Ala	
130 135 140	
gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg	480
Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp	
145 150 155 160	
tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta	528
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val	
165 170 175	
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg	576

Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Met	
			180					185					190			
att	att	gtt	gtt	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggt	624
Ile	Ile	Val	Val	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly	
		195					200					205				
tac	cta	atg	ggg	ggc	gaa	ggg	gta	tac	gct	tca	aac	tta	aac	ctc	ata	672
Tyr	Leu	Met	Gly	Gly	Glu	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile	
	210					215					220					
tat	aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggg	ttg	atc	att	720
Tyr	Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	
225					230					235					240	
tgg	aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct						753
Trp	Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala						
				245					250							

<210> 55

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 55

Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Ala	Ile	Ala	Leu	Pro	Ser	
1				5					10					15		
Phe	Ala	Ala	Ala	Gly	Gly	Asp	Leu	Asp	Ile	Ser	Asp	Thr	Val	Gly	Val	
			20					25					30			
Ser	Phe	Trp	Leu	Val	Thr	Ala	Gly	Met	Leu	Ala	Ala	Thr	Val	Phe	Phe	
	35						40					45				
Phe	Val	Glu	Arg	Asp	Gln	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr	
	50					55					60					

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Ser Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 56

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalB6; GenBank # AF349998

<400> 56

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	
ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	
tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	
aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat	288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu	

100	105	110	
att ctt gct gct tgt aca aat gtt gct gct tca tta ttt aag aag ctt			384
Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu			
115	120	125	
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct			432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala			
130	135	140	
gga tta gct cct gta tgg cct gct ttc att att ggt atg gct gga tgg			480
Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp			
145	150	155	160
tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta			528
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val			
165	170	175	
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg gtg			576
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val			
180	185	190	
att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt			624
Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly			
195	200	205	
tac cta atg ggt ggc gaa ggt gta tac gct tca aac cta aac ctt ata			672
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile			
210	215	220	
tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att			720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile			
225	230	235	240
tgg aat gtt gct gtt aaa gaa tct tct aat gct			753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala			
245	250		

<210> 57

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 57

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val
180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 58

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacteria

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalB8; GenBank #AF350000

<400> 58

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe 35 40 45	144
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr 50 55 60	192
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met 65 70 75 80	240
aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr 85 90 95	288
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu 100 105 110	336
att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu 115 120 125	384
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala 130 135 140	432
gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp 145 150 155 160	480
tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val 165 170 175	528
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met 180 185 190	576
att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly 195 200 205	624
tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile 210 215 220	672

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalE1;GenBank# AF350001

<400> 60

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	

gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	

aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat	288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr	
85 90 95	

att gat tgg tta tta act gtt cca tta caa gtg gtt gag ttc tat cta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu	
100 105 110	

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt	384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu	
115 120 125	

cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct	432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala	
130 135 140	

gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg	480
---	-----

Gly	Leu	Ala	Pro	Val	Leu	Pro	Ala	Phe	Ile	Ile	Gly	Met	Ala	Gly	Trp	
145					150					155					160	
tta	tac	atg	att	tat	gag	cta	tat	atg	ggg	gaa	ggc	aag	gct	gct	gta	528
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val	
				165					170					175		
agt	act	gca	agt	cct	gct	gtt	aac	cct	gca	tac	aac	gca	atg	atg	atg	576
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Pro	Ala	Tyr	Asn	Ala	Met	Met	Met	
			180					185					190			
att	att	gtt	gtt	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggg	624
Ile	Ile	Val	Val	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly	
		195					200					205				
tac	cta	atg	ggg	ggc	gaa	ggg	gta	tac	gct	tca	aac	tta	aac	ctt	ata	672
Tyr	Leu	Met	Gly	Gly	Glu	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile	
	210					215					220					
tat	aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggg	ttg	atc	att	720
Tyr	Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	
225					230					235					240	
tgg	aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct						753
Trp	Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala						
				245					250							
<210>	61															
<211>	251															
<212>	PRT															
<213>	Naturally occurring gamma proteobacteria															
<400>	61															
Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Ala	Ile	Ala	Leu	Pro	Ser	
1				5					10					15		
Phe	Ala	Ala	Ala	Gly	Gly	Asp	Leu	Asp	Ile	Ser	Asp	Thr	Val	Gly	Val	
			20					25					30			

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Pro Ala Tyr Asn Ala Met Met Met
 180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 62

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalE6; GenBank#AF350002

<400> 62

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gta ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	

gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	

aga	ggt	ggt	tgg	ata	gat	act	ggt	gat	aca	cca	aca	gta	ttt	aga	tat		288
Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr		
			85						90					95			
att	gat	tgg	tta	tta	act	ggt	cca	tta	caa	atg	ggt	gag	ttc	tat	cta		336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gct	gct	tgt	aca	agt	ggt	gct	gct	tca	tta	ttt	aag	aag	ctt		384
Ile	Leu	Ala	Ala	Cys	Thr	Ser	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu		
			115				120					125					
cta	gct	ggt	tca	tta	gta	atg	tta	ggt	gct	gga	ttt	gca	ggc	gaa	gct		432
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Phe	Ala	Gly	Glu	Ala		
			130			135					140						
ggt	tta	gct	cct	gta	tta	cct	gct	ttc	att	att	ggt	atg	gct	gga	tgg		480
Gly	Leu	Ala	Pro	Val	Leu	Pro	Ala	Phe	Ile	Ile	Gly	Met	Ala	Gly	Trp		
145				150					155						160		
tta	tac	atg	att	tat	gag	cta	cat	atg	ggt	gaa	ggt	aag	gct	gct	gta		528
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	His	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val		
			165						170				175				
agt	act	gca	agt	cct	gct	ggt	aac	tct	gca	tac	aac	gca	atg	atg	aag		576
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Lys		
			180				185					190					
att	att	ggt	att	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggt		624
Ile	Ile	Val	Ile	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly		
			195			200						205					
tac	cta	atg	agt	ggt	gac	ggt	gta	tac	gct	tca	aac	tta	aac	ctt	ata		672
Tyr	Leu	Met	Ser	Gly	Asp	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile		
			210			215					220						
tat	aac	ctt	gct	gac	ttt	ggt	aac	aag	att	cta	ttt	ggt	ttg	atc	att		720
Tyr	Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile		
225				230			235								240		
tgg	aat	ggt	gct	ggt	aaa	gaa	tct	tct	aat	gct							753
Trp	Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala							
				245					250								

<210> 63

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 63

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys
180 185 190

Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 64

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalE7; GenBank# AF350003

<400> 64

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca

48

Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Ala	Ile	Ala	Leu	Pro	Ser	
1				5					10					15		
ttt	gct	gct	gct	ggt	ggc	gat	cta	gat	ata	agt	gat	act	gtt	ggt	ggt	96
Phe	Ala	Ala	Ala	Gly	Gly	Asp	Leu	Asp	Ile	Ser	Asp	Thr	Val	Gly	Val	
			20					25					30			
tca	ttc	tgg	ctg	ggt	aca	gct	ggt	atg	tta	gcg	gca	act	gtg	ttc	ttt	144
Ser	Phe	Trp	Leu	Val	Thr	Ala	Gly	Met	Leu	Ala	Ala	Thr	Val	Phe	Phe	
		35					40					45				
ttt	gta	gaa	aga	gac	caa	gtc	agc	gct	aag	tgg	aaa	act	tca	ctt	act	192
Phe	Val	Glu	Arg	Asp	Gln	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr	
	50					55					60					
gta	tct	ggt	tta	att	act	ggt	ata	gct	ttt	tgg	cat	tat	ctc	tat	atg	240
Val	Ser	Gly	Leu	Ile	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Leu	Tyr	Met	
65					70					75					80	
aga	ggt	ggt	tgg	ata	gat	act	ggt	gat	acc	cca	aca	gta	ttc	aga	tat	288
Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr	
				85					90					95		
att	gat	tgg	tta	tta	act	ggt	cca	tta	caa	atg	ggt	gag	ttc	tat	cta	336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu	
			100					105					110			
att	ctt	gct	gct	tgt	aca	agt	ggt	gct	gct	tca	tta	ttt	aag	aag	ctt	384
Ile	Leu	Ala	Ala	Cys	Thr	Ser	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu	
		115					120					125				
cta	gct	ggt	tca	tta	gta	atg	tta	ggt	gct	gga	ttt	gca	ggc	gaa	gct	432
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Phe	Ala	Gly	Glu	Ala	
	130					135					140					
gga	tta	gct	cct	gta	tta	cct	gct	ttc	att	att	ggt	atg	gct	gga	tgg	480
Gly	Leu	Ala	Pro	Val	Leu	Pro	Ala	Phe	Ile	Ile	Gly	Met	Ala	Gly	Trp	
145					150					155					160	
cta	tac	atg	att	tat	gag	cta	tat	atg	ggt	gaa	ggt	aag	gct	gct	gta	528
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val	
				165					170					175		
agt	act	gca	agt	cct	gct	ggt	aac	tct	gca	tac	aac	gca	atg	atg	atg	576
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Met	
			180					185					190			
att	att	ggt	ggt	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggt	624

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

tac cta atg ggt ggc gaa ggc gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 65

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 65

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250